# 93870, A HUMAN G-PROTEIN COUPLED RECEPTOR AND USES THEREFOR

## Cross-References to Related Applications

**[0001]** This application claims the benefit of U.S. Provisional Application Number 60/272,677 filed March 1, 2001, the contents of which are incorporated herein by this reference.

## Background of the Invention

**[0002]** G-protein coupled receptors (GPCRs) are seven transmembrane domain proteins that mediate signal transduction of a diverse number of ligands through heterotrimeric G proteins (Strader, C. D. et al. (1994) *Annu. Rev. Biochem.* 63: 101-132). G protein-coupled receptors (GPCRs), along with G-proteins and effector proteins (e.g., intracellular enzymes and channels), are the components of a modular signaling system. Upon ligand binding to an extracellular portion of a GPCR, different G proteins are activated, which in turn modulate the activity of different intracellular effector enzymes and ion channels (Gutkind, J.S. (1998) *J. Biol. Chem.* 273: 1839-1842; Selbie, L.A. and Hill, S.J. (1998) *Trends Pharmacol. Sci.* 19:87-93).

**[0003]** G proteins represent a family of heterotrimeric proteins composed of α, β and γ subunits, which bind guanine nucleotides. These proteins are usually linked to cell surface receptors (e.g., a GPCR). Following ligand binding to a GPCR, a conformational change is transmitted to the G protein, which causes the α-subunit to exchange a bound GDP molecule for a GTP molecule and to dissociate from the βγ-subunits. The GTP-bound form of the α-subunit typically functions as an effector-modulating moiety, leading to the production of second messengers, such as cyclic AMP (e.g., by activation of adenylate cyclase), diacylglycerol or inositol phosphates. Greater than 20 different types of α-subunits are known in man, which associate with a smaller pool of β and γ subunits. Examples of mammalian G proteins include $G_i$, $G_o$, $G_q$, $G_s$ and $G_t$ (Lodish H. et al. Molecular Cell Biology, (Scientific American Books Inc., New York, N.Y., 1995).

**[0004]** GPCRs are of critical importance to several systems including the endocrine system, the central nervous system and peripheral physiological processes. The GPCR genes and gene-products are also believed to be causative agents of disease (Spiegel et al. (1993) *J.*

10085233.022802

*Clin. Invest.* 92:1119-1125); McKusick and Amberger (1993) *J. Med. Genet.* 30:1-26). Given the important biological roles and properties of GPCRs, there exists a need for the identification of novel genes encoding such proteins as well as for the discovery of modulators of such molecules for use in regulating a variety of normal and/or pathological cellular processes.

## Summary of the Invention

**[0005]** The present invention is based, in part, on the discovery of a novel GPCR, with similarities to Subfamily I of G- protein coupled receptor type proteins (GPCRs), referred to herein as "93870". The nucleotide sequence of a cDNA encoding 93870 is shown in SEQ ID NO:1, and the amino acid sequence of a 93870 polypeptide is shown in SEQ ID NO:2. In addition, the nucleotide sequence of the coding region is depicted in SEQ ID NO:3.

**[0006]** Accordingly, in one aspect, the invention features a nucleic acid molecule which encodes a 93870 protein or polypeptide, e.g., a biologically active portion of the 93870 protein. In a preferred embodiment, the isolated nucleic acid molecule encodes a polypeptide having the amino acid sequence of SEQ ID NO:2. In other embodiments, the invention provides an isolated 93870 nucleic acid molecule having the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, and the sequence of the DNA insert of the plasmid deposited with ATCC on ___ as Accession Number ___ (hereafter, "the deposited nucleotide sequence").

**[0007]** In still other embodiments, the invention provides nucleic acid molecules that are sufficiently or substantially identical (e.g., naturally occurring allelic variants) to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, and the deposited nucleic acid sequence. In other embodiments, the invention provides a nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of one of SEQ ID NO:1, SEQ ID NO:3, and the deposited nucleotide sequence, wherein the nucleic acid encodes a full length 93870 protein or an active fragment thereof.

**[0008]** In a related aspect, the invention further provides nucleic acid constructs which include a 93870 nucleic acid molecule described herein. In certain embodiments, the nucleic acid molecules of the invention are operatively linked to native or heterologous regulatory sequences. Also included, are vectors and host cells containing the 93870 nucleic acid molecules of the invention e.g., vectors and host cells suitable for producing polypeptides.

[0009] In another related aspect, the invention provides nucleic acid fragments suitable as primers or hybridization probes for the detection of 93870-encoding nucleic acids.

[0010] In still another related aspect, isolated nucleic acid molecules that are antisense to a 93870 encoding nucleic acid molecule are provided.

[0011] In another aspect, the invention features 93870 polypeptides, and biologically active or antigenic fragments thereof that are useful, e.g., as reagents or targets in assays applicable to treatment and diagnosis of 93870-mediated or related disorders, e.g., GPCR disorders as described herein. In another embodiment, the invention provides 93870 polypeptides having a 93870 activity. Preferred polypeptides are 93870 proteins including at least one, two, three, four, five, six or seven transmembrane domains, and, preferably, having a 93870 activity, e.g., a 93870 activity as described herein. Preferred polypeptides are 93870 proteins including at least one receptor coupled G-protein transmembrane domain.

10085233.022802

[0012] In other embodiments, the invention provides 93870 polypeptides, *e.g.*, a 93870 polypeptide having the amino acid sequence shown in SEQ ID NO:2 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC Accession Number ____; an amino acid sequence that is sufficiently or substantially identical to the amino acid sequence shown in SEQ ID NO:2 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC Accession Number ____; or an amino acid sequence encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3 or the nucleotide sequence of the insert of the plasmid deposited with ATCC Accession Number ____, wherein the nucleic acid encodes a full length 93870 protein or an active fragment thereof.

[0013] In a related aspect, the invention further provides nucleic acid constructs which include a 93870 nucleic acid molecule described herein.

[0014] In a related aspect, the invention provides 93870 polypeptides or fragments operatively linked to non-93870 polypeptides to form fusion proteins.

[0015] In another aspect, the invention features antibodies and antigen-binding fragments thereof, that react with, or more preferably specifically or selectively bind 93870 polypeptides.

[0016] In another aspect, the invention provides methods of screening for compounds that modulate the expression or activity of the 93870 polypeptides or nucleic acids.

**[0017]** In still another aspect, the invention provides a process for modulating 93870 polypeptide or nucleic acid expression or activity, e.g. using the screened compounds. In certain embodiments, the methods involve treatment of conditions related to aberrant activity or expression of the 93870 polypeptides or nucleic acids, such as conditions involving aberrant or deficient cellular proliferation or differentiation.

**[0018]** The invention also provides assays for determining the activity of or the presence or absence of 93870 polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

**[0019]** In another aspect, the invention features a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, *e.g.,* a nucleic acid or peptide sequence. At least one address of the plurality has a capture probe that recognizes a 93870 molecule. In one embodiment, the capture probe is a nucleic acid, *e.g.*, a probe complementary to a 93870 nucleic acid sequence. In another embodiment, the capture probe is a polypeptide, *e.g.*, an antibody specific for 93870 polypeptides. Also featured is a method of analyzing a sample by contacting the sample to the aforementioned array and detecting binding of the sample to the array.

**[0020]** In further aspect the invention provides assays for determining the presence or absence of a genetic alteration in a 93870 polypeptide or nucleic acid molecule, including for disease diagnosis.

10085233.022802

**Brief Description of the Drawing**

**[0021]** *Figure 1* depicts a hydropathy plot of human 93870. Relatively hydrophobic residues are shown above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) are indicated by short vertical lines below the hydropathy trace. The numbers corresponding to the amino acid sequence of human 93870 are indicated on the x axis. Polypeptides of the invention include fragments which include: all or part of a hydrophobic sequence, *e.g.*, a sequence above the dashed line, *e.g.*, the sequence from about amino acid 60 to 80, from about 140 to 160, and from about 210 to 230 of SEQ ID NO:2; all or part of a hydrophilic sequence, *e.g.*, a sequence below the dashed line, *e.g.*, the sequence from about amino acid 125 to 135, from about 160 to 170, and from about 210 to 230 of SEQ ID NO:2; a sequence which includes a Cys, or a glycosylation site.

## Detailed Description of the Invention

**[0022]** The human 93870 sequence (SEQ ID NO:1), which is approximately 1684 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 942 nucleotides, including the termination codon (nucleotides indicated as coding of SEQ ID NO:1; SEQ ID NO:3). The coding sequence encodes a 313 amino acid protein (SEQ ID NO:2).

**[0023]** A plasmid containing the nucleotide sequence encoding human 93870 was deposited with American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Number ______. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

**[0024]** The 93870 protein contains a number of structural characteristics in common with members of the G-protein coupled receptor family, and in particular the subfamily-1 of the G-protein coupled receptors. The term "family" when referring to the protein and nucleic acid molecules of the invention means two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin as well as other distinct proteins of human origin, or alternatively, can contain homologues of non-human origin, e.g., rat or mouse proteins. Members of a family can also have common functional characteristics.

**[0025]** The G-protein coupled receptor family of seven transmembrane proteins is an extensive group of proteins, which transduce extracellular signals triggered by, e.g., hormones, neurotransmitters, odorants and light, by interaction with guanine nucleotide-binding (G) proteins. The N-terminus of G-protein coupled receptors is typically located on the extracellular side of the membrane and is often glycosylated, while the C-terminus is cytoplasmic and generally phosphorylated. G-protein coupled receptors typically have seven hydrophobic membrane spanning regions. Three extracellular loops alternate with three intracellular loops to link the seven transmembrane regions. Some G-protein coupled receptors possess a signal peptide. Generally, the most conserved portions of G-protein

coupled receptors are the transmembrane regions and the first two cytoplasmic loops. A conserved arginine-aromatic doublet is present in the N-terminal extremity of the second cytoplasmic loop and may be implicated in the interaction with G proteins. Alignments of the domains of 1308 representative GPCRs can be found at http://www.toulouse.inra.fr/prodom/cgi-bin/ReqProdomII.pl?id_dom0=PD000009&prodom_release=2000.1. Thus, the 93870 proteins of the invention include the following structural features that demonstrate their inclusion in the G-protein coupled receptor family: (1) an N-terminal exracellular domain; (2) seven transmembrane domains; (3) three extracellular loops; (4) three cytoplasmic loops, one of which includes the conserved arginine-aromatic doublet; and (5) a C-terminal cytoplasmic domain.

**[0026]** In one embodiment, a 93870 protein includes at least one extracellular domain. When located at the N-terminal domain the extracellular domain is referred to herein as an "N-terminal extracellular domain" in the amino acid sequence of the protein. As used herein, an "N-terminal extracellular domain" includes an amino acid sequence having about 1-100, preferably about 1-75, more preferably about 1-50, even more preferably about 1-30 amino acid residues in length and is located outside of a cell or extracellularly. The C-terminal amino acid residue of a "N-terminal extracellular domain" is adjacent to an N-terminal amino acid residue of a transmembrane domain in a naturally-occurring 93870 or 93870-like protein. For example, an N-terminal extracellular domain is located at about amino acid residues 1-27 of SEQ ID NO:2.

**[0027]** In a preferred embodiment, a 93870 polypeptide or protein has an “N-terminal extracellular domain” or a region which includes at least about 1-100, preferably about 1-50, and even more preferably about 1-30 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with an “N-terminal extracellular domain,” e.g., the N-terminal extracellular domain of human 93870 (e.g., residues 1-27 of SEQ ID NO:2). Preferably, the N-terminal extracellular domain is capable of interacting with (e.g., binding to) an extracellular signal, for example, a ligand or a cell surface receptor. Most preferably, the N-terminal extracellular domain mediates protein-protein interactions, signal transduction and/or cell adhesion.

**[0028]** In another embodiment, a 93870 protein includes at least one, two, three, four, five, six, or preferably, seven transmembrane domains. As used herein, the term "transmembrane domain" includes an amino acid sequence of about 15 amino acid residues

in length that spans the plasma membrane. More preferably, a transmembrane domain includes about at least 20, 23, 24, 25, 30 or 35 amino acid residues and spans the plasma membrane. Transmembrane domains are rich in hydrophobic residues, and typically have an α-helical structure. In a preferred embodiment, at least 50%, 60%, 70%, 80%, 90%, 95% or more of the amino acids of a transmembrane domain are hydrophobic, e.g., leucines, isoleucines, tyrosines, or tryptophans. Transmembrane domains are described in, for example, htto://pfam.wustl.edu/cgi-bin/getdesc?name=7tm-1, and Zagotta W.N. et al, (1996) *Annual Rev. Neuronsci.* 19: 235-63, the contents of which are incorporated herein by reference. The human 93870 also has seven transmembrane domains extending from about amino acid 28 (extracellular end) to about amino acid 51 (cytoplasmic end) of SEQ ID NO:2; from about amino acid 58 (cytoplasmic end) to about amino acid 79 (extracellular end) of SEQ ID NO:2; from about amino acid 98 (extracellular end) to about amino acid 119 (cytoplasmic end) of SEQ ID NO:2; from about amino acid 140 (cytoplasmic end) to about amino acid 160 (extracellular end) of SEQ ID NO:2; from about amino acid 192 (extracellular end) to about amino acid 216 (cytoplasmic end) of SEQ ID NO:2; from about amino acid 236 (cytoplasmic end) to about amino acid 252 (extracellular end) of SEQ ID NO:2; from about amino acid 276 (extracellular end) to about amino acid 299 (cytoplasmic end) of SEQ ID NO:2;

**[0029]** In a preferred embodiment, a 93870 polypeptide or protein has at least one transmembrane domain or a region which includes at least 15, 20, 23, 24, 25, 30 or 35 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a "transmembrane domain," e.g., at least one transmembrane domain of human 93870 (e.g., residues 28-51, 58-79, 98-119, 140-160, 192-216, 236-252, and 276-299 of SEQ ID NO:2). Preferably, the transmembrane domain is involved in transducing a signal, e.g., an extracellular signal across a cell membrane, and/or activates a signal transduction pathway.

**[0030]** In another embodiment, a 93870 protein includes at least one extracellular loop. As defined herein, the term "loop" includes an amino acid sequence having a length of at least about 4, preferably about 5-10, more preferably about 10-20, and even more preferably about 20-30 amino acid residues, and which connects two transmembrane domains within a protein or polypeptide. Accordingly, the N-terminal amino acid of a loop is adjacent to a C-terminal amino acid of a transmembrane domain in a naturally-occurring 93870 or 93870-like molecule, and the C-terminal amino acid of a loop is adjacent to an N-terminal amino acid of a transmembrane domain in a naturally-occurring 93870 or 93870-like molecule. As

used herein, an "extracellular loop" includes an amino acid sequence located outside of a cell, or extracellularly. For example, an extracellular loop can be found at about amino acids 80-97, 161-191, and 253-275 of SEQ ID NO:2.

**[0031]** In a preferred embodiment, a 93870 polypeptide or protein has at least one extracellular loop or a region which includes at least about 4, preferably about 5-10, more preferably about 10-20, more preferably about 20-30, and most preferably about 30-40 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with an "extracellular loop," e.g., at least one extracellular loop of human 93870 (e.g., residues 80-97, 161-191, and 253-275 of SEQ ID NO:2).

**[0032]** In another embodiment, a 93870 protein includes at least one cytoplasmic loop, also referred to herein as a cytoplasmic domain. As used herein, a "cytoplasmic loop" includes an amino acid sequence having a length of at least about 4, preferably about 5-10, more preferably about 10-20, more preferably about 20-30, and most preferably about 30-40 amino acid residues located within a cell or within the cytoplasm of a cell. For example, a cytoplasmic loop is found at about amino acids 52-57, 120-139, and 217-235 of SEQ ID NO:2.

**[0033]** In a preferred embodiment, a 93870 polypeptide or protein has at least one cytoplasmic loop or a region which includes at least about 4, preferably about 5-10, more preferably about 10-20, more preferably about 20-30, and most preferably about 30-40 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with an "cytoplasmic loop," e.g., at least one cytoplasmic loop of human 93870 (e.g., residues 52-57, 120-139, and 217-235 of SEQ ID NO:2). Within the cytoplasmic loop, the human 93870 polypeptide also has a arginine, tyrosine (aromatic) doublet at amino acids about 121-122 of SEQ ID NO:2 respectively, matching the conserved arginine-aromatic doublet that is present in the N-terminal extremity of the second cytoplasmic loop of most G protein coupled transmembrane receptors. The conserved doublet is implicated in the interaction with the G-protein.

**[0034]** In another embodiment, a 93870 protein includes a "C-terminal cytoplasmic domain", also referred to herein as a C-terminal cytoplasmic tail, in the sequence of the protein. As used herein, a "C-terminal cytoplasmic domain" includes an amino acid sequence having a length of at least about 5, and preferably about 10-50 amino acid residues, and is located within a cell or within the cytoplasm of a cell. Accordingly, the N-terminal amino acid residue of a "C-terminal cytoplasmic domain" is adjacent to a C-

10085233.022802

terminal amino acid residue of a transmembrane domain in a naturally-occurring 93870 or 93870-like protein. For example, a C-terminal cytoplasmic domain is found at about amino acid residues 300-313 of SEQ ID NO:2.

**[0035]** In a preferred embodiment, a 93870 polypeptide or protein has a C-terminal cytoplasmic domain or a region which includes at least about 5, and preferably about 10-50 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with an "C-terminal cytoplasmic domain," e.g., the C-terminal cytoplasmic domain of human 93870 (e.g., residues 300-313 of SEQ ID NO:2).

**[0036]** Based on structural similarities, members of the GPCR family have been classified into various subfamilies, including: Subfamily I which comprises receptors typified by rhodopsin and the beta2-adrenergic receptor and currently contains over 200 unique members (reviewed by Dohlman *et al.* (1991) *Annu. Rev. Biochem.* 60:653-688); Subfamily II, which includes the parathyroid hormone/calcitonin/secretin receptor family (Juppner *et al.* (1991) *Science* 254:1024-1026; Lin *et al.* (1991) *Science* 254:1022-1024); Subfamily III, which includes the metabotropic glutamate receptor family in mammals, such as the GABA receptors (Nakanishi *et al.* (1992) *Science* 258: 597-603); Subfamily IV, which includes the cAMP receptor family that is known to mediate the chemotaxis and development of *D. discoideum* (Klein *et al.* (1988) *Science* 241:1467-1472); and Subfamily V, which includes the fungal mating pheromone receptors such as STE2 (reviewed by Kurjan I *et al.* (1992) *Annu. Rev. Biochem.* 61:1097-1129). Within each family, distinct, highly conserved motifs have been identified. These motifs have been suggested to be critical for the structural integrity of the receptor, as well as for coupling to G proteins.

10095233.022802

**[0037]** The most conserved parts of these GPCR Subfamily I proteins are the transmembrane regions and the first two cytoplasmic loops. A conserved arginine-aromatic doublet is present in the N-terminal extremity of the second cytoplasmic loop and is implicated in the interaction with G proteins.

**[0038]** The human 93870 polypeptide also contains a receptor coupled G-protein receptor transmembrane domain derived from ProDom PD000009 (Release 2000.1; http//protein.toulouse.inra.fr/prodom/cgi-bin/NewBlastProdomII.) located at about amino acid residues 56-123 of SEQ ID NO:2. To identify the presence of a "receptor coupled G-protein transmembrane " domain in a 93870 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be aligned against a database of 1308 family members (using BLASTP

2.0a19MP-WashU (05 Feb-1998)). Sequences containing high scoring segment pairs are reported from ProDom. As used herein, the term "receptor coupled G-protein transmembrane domain" includes an amino acid sequence of about 25-125 amino acid residues in length, preferably about 50-100 amino acids, more preferably about 60-80 amino acids, or about 67 amino acids and has a bit score for the alignment of the sequence to the receptor coupled G-protein transmembrane domain of at least about 50, preferably about 100, or more preferably about 125 or greater, and an E-value of about 3.7e-6 or less, more preferably about 3.7e-7 or less, and most preferably about 3.7-8 or less.

**[0039]** For further identification of a 93870 protein sequence as a G-protein coupled receptor, the amino acid sequence of the protein was searched against a database of domains, *e.g.*, the ProDom database (Corpet *et al.* (1999), *Nucl. Acids Res.* 27:263-267). The ProDom protein domain database consists of an automatic compilation of homologous domains. Current versions of ProDom are built using recursive PSI-BLAST searches (Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402; Gouzy *et al.* (1999) *Computers and Chemistry* 23:333-340) of the SWISS-PROT 38 and TREMBL protein databases. The database automatically generates a consensus sequence for each domain. A BLAST search was performed against the HMM database yielding a consensus amino acid sequence for a GPCR derived from ProDom PD000009 (Release 2000.1; http//protein.toulouse.inra.fr/prodom/cgi-bin/NewBlastProdomII.). The consensus amino acid sequence from PD000009 is amino acids 56 to 123 of a murine GPCR, SWISS-PROT:P51676 (SEQ ID NO:5). For additional identification of the 93870 protein sequence as a GPCR, the full murine amino acid sequence for the GPCR derived from ProDom PD000009 (Release 2000.1; http//protein.toulouse.inra.fr/prodom/cgi-bin/NewBlastProdomII) is depicted as SEQ ID NO:4 and represents amino acids 1 to 356 of SWISS-PROT:P51676. An alignment of the 93870 protein with SWISS-PROT:P51676 (SEQ ID NO:5) demonstrates about 21% sequence identity between the two sequences (as calculated in matblas from the blosum62.iij matrix).

**[0040]** The receptor coupled G-protein transmembrane domain has been assigned the ProDom Accession Number PD000009 (Release 2000.1; http//protein.toulouse.inra.fr/prodom/cgi-bin/NewBlastProdomII.). The receptor coupled G-protein transmembrane domain in 93870 as predicted by ProDom has a bit score of 141, and the E-value of 3.7e-9. A top scoring alignment of the receptor coupled G-protein

10085233.022802

transmembrane domain (amino acids 56 to 123 of SEQ ID NO:2) of human 93870 against a murine amino acid sequence is depicted as (SEQ ID NO:5).

**[0041]** The receptor coupled G- protein domain typically has the following consensus sequence:

**[0042]** [GSTALIVMFYWC]-[GSTANCPDE-{EDPKRH}-x(2)-[LIVMNQGA]-x(2)-[LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)-[LIVM] (SEQ ID NO:6).

**[0043]** This consensus sequence is located at amino acid residues 109 to 125 of SEQ ID NO:2. In this consensus sequence pattern, each element in the pattern is separated by a dash (-); square [ ] brackets indicate the particular residues that are accepted at that position; x indicates any residue is accepted at that position; a whole number in parenthesis following an x indicates any amino acid repetition of a particular element is accepted for that specified number of residues i.e. x(2); { } brackets indicate that the particular amino acid in that position can be any except those enclosed in the bracket.

**[0044]** The 93870 polypeptide of the invention contains a portion of the receptor coupled G-protein domain representing 76% of the receptor coupled G-protein receptor consensus pattern described in PROSITE pattern PDOC00210 (http://www.expasy.ch/cgi-bin/nicedoc.pl?PDOC00210). The receptor coupled G-protein consensus pattern of the 93870 polypeptide differs at amino acid residues 114, 118, 120 and 125 of SEQ ID NO:2, wherein a "Y" is substituted for any of the " LIVMNQGA " residues at amino acid residue 114 of SEQ ID NO:2, a "L" is substituted for any of the "GSTANC" residues at amino acid residue 118 of SEQ ID NO:2, a "T" is substituted for any of the "DENH" residues at amino acid residue 120 of SEQ ID NO:2, and a "F" is substituted for any of the " LIVM " residues at amino acid residue 125 of SEQ ID NO:2.

**[0045]** The human 93870 protein has two N-glycosylation sites (Prosite Pattern Number PS00001) at about amino acid residues 13-16 and 17-20 of SEQ ID NO:2; two protein kinase C phosphorylation sites (Prosite Pattern Number PS00005) at about amino acid residues 14-16 and 267-269 of SEQ ID NO:2; three casein kinase II phosphorylation sites (Prosite Pattern Number PS00006) located at about amino acid residues 18-21, 225-228, and 284-287 of SEQ ID NO:2; one tyrosine kinase phosphorylation site (Prosite Pattern Number PS00007) at about amino acid residues 176-178 of SEQ ID NO:2; three N-myristylation sites (Prosite Pattern Number PS00008) at about amino acid residues 38-43, 92-97, and 305-310 of SEQ ID NO:2.



10085233.022802

**[0046]** General information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers can be found at Sonnhammer *et al.* (1997*) Protein* 28:405-420 and http://www.psc.edu/general/ software/packages/pfam/pfam.html.

**[0047]** Accordingly, in one embodiment of the invention, a 93870 includes at least one, two or three, preferably four, five, or six, and most preferably seven, transmembrane domains and/or at least one, two, preferably three cytoplasmic loops, and/or at least one, two, or preferably three extracellular loops. In another embodiment, the 93870 further includes an N-terminal extracellular domain and/or a C-terminal cytoplasmic domain. In another embodiment, the 93870 can include seven transmembrane domains, three cytoplasmic loops, three extracellular loops and can further include an N-terminal extracellular domain and/or a C-terminal cytoplasmic domain.

**[0048]** The 93870 molecules of the present invention can further include at least one N-glycosylation site. The 93870 molecule can have one, preferably two protein kinase C phosphorylation sites. The 93870 molecules can include at least one, two, preferably three casein kinase II phosphorylation sites. The 93870 molecules can have at least one tyrosine kinase phophorylation site. The 93870 molecules can additionally include at least one, two, preferably three N-myristylation sites.

10085233.022802

**[0049]** As the 93870 polypeptides of the invention may modulate 93870 activities, they may be useful for developing novel diagnostic and therapeutic agents for 93870-mediated or related disorders, as described below.

**[0050]** As used herein, a "93870 activity", "biological activity of 93870" or "functional activity of 93870", refers to an activity exerted by a 93870 protein, polypeptide or nucleic acid molecule on e.g., a 93870-responsive cell or on a 93870 substrate, e.g., a protein substrate, as determined *in vivo* or *in vitro*. In one embodiment, a 93870 activity is a direct activity, such as an association with a 93870 target molecule. A "target molecule" or "binding partner" is a molecule with which a 93870 protein binds or interacts in nature. In an exemplary embodiment, is a 93870 ligand. A 93870 activity can also be an indirect activity, e.g., a cellular signaling activity mediated by interaction of the 93870 protein with a 93870 ligand.

**[0051]** Based on the above-described sequence similarities, the 93870 molecules of the present invention are predicted to have similar biological activities as those of the G-protein coupled receptor Subfamily I members. For example, the 93870 protein of the present invention is has one more of the following activities: (1) the ability to regulate, sense,

and/or transmitting an extracellular signal into a cell, for example, an osteoblast, neutrophil, megakaryocyte, or bone marrow mononuclear cell; (2) the ability to interact with (e.g., bind to) an extracellular signal or a cell surface receptor; (3) the ability to mobilize an intracellular molecule that participates in a signal transduction pathway (e.g., adenylate cyclase or phosphatidylinositol 4,5-bisphosphate ($PIP_2$), inositol 1,4,5-triphosphate ($IP_3$)); (5) the ability to control production or secretion of molecules; (6) the ability to alter the structure of a cellular component; (7) the ability to modulate cell proliferation, e.g., synthesis of DNA; and (8) the ability to modulate cell growth, migration, differentiation, survival, and/or function e.g., of the cells or tissue in which they are expressed (e.g., bone marrow mononuclear cells, (e.g., granulocytes (e.g., neutrophils), osteoblasts, and megakaryocytes). (9) the ability to modulate the synthesis of organic components of bone matrix (type I collagen, proteoglycans, and glycoproteins) by osteoblasts; (10) the ability to sense and mediate cellular response to environmental stimuli, (e.g., small molecules, protein ligands); (11) the ability to sense and mediate cellular response to biological messengers, (e.g., secreted hormones); and (12) the ability to signal through G proteins. Thus, the 93870 molecules can act as novel diagnostic targets and therapeutic agents for controlling GPCR associated disorders.

**[0052]** The response mediated by a 93870 receptor protein depends on the type of cell in which it is expressed. For example, in some cells, binding of a ligand to the receptor protein may stimulate an activity such as release of compounds, gating of a channel, cellular adhesion, migration, differentiation, etc., through phosphatidylinositol or cyclic AMP metabolism and turnover while in other cells, the binding of the ligand will produce a different result. Regardless of the cellular activity/response modulated by the receptor protein, it is universal that the protein is a GPCR and interacts with G proteins to produce one or more secondary signals, in a variety of intracellular signal transduction pathways, e.g., through phosphatidylinositol or cyclic AMP metabolism and turnover, in a cell. As used herein, a "signaling transduction pathway" refers to the modulation (e.g., stimulation or inhibition) of a cellular function/activity upon the binding of a ligand to the GPCR (93870 protein). Examples of such functions include mobilization of intracellular molecules that participate in a signal transduction pathway, e.g., phosphatidylinositol 4,5-bisphosphate ($PIP_2$), inositol 1,4,5-triphosphate ($IP_3$) and adenylate cyclase.

**[0053]** As used herein, "phosphatidylinositol turnover and metabolism" refers to the molecules involved in the turnover and metabolism of phosphatidylinositol 4,5-bisphosphate

($PIP_2$) as well as to the activities of these molecules. $PIP_2$ is a phospholipid found in the cytosolic leaflet of the plasma membrane. Binding of ligand to the receptor activates, in some cells, the plasma-membrane enzyme phospholipase C that in turn can hydrolyze $PIP_2$ to produce 1,2-diacylglycerol (DAG) and inositol 1,4,5-triphosphate ($IP_3$). Once formed $IP_3$ can diffuse to the endoplasmic reticulum surface where it can bind an $IP_3$ receptor, e.g., a calcium channel protein containing an $IP_3$ binding site. $IP_3$ binding can induce opening of the channel, allowing calcium ions to be released into the cytoplasm. $IP_3$ can also be phosphorylated by a specific kinase to form inositol 1,3,4,5-tetraphosphate ($IP_4$), a molecule which can cause calcium entry into the cytoplasm from the extracellular medium. $IP_3$ and $IP_4$ can subsequently be hydrolyzed very rapidly to the inactive products inositol 1,4-biphosphate ($IP_2$) and inositol 1,3,4-triphosphate, respectively. These inactive products can be recycled by the cell and used to synthesize $PIP_2$. The other second messenger produced by the hydrolysis of $PIP_2$, namely 1,2-diacylglycerol (DAG), remains in the cell membrane where it can serve to activate the enzyme protein kinase C. Protein kinase C is usually found soluble in the cytoplasm of the cell, but upon an increase in the intracellular calcium concentration, this enzyme can move to the plasma membrane where it may be activated by DAG. The activation of protein kinase C in different cells results in various cellular responses such as the phosphorylation of glycogen synthase, or the phosphorylation of various transcription factors, e.g., NF-κB. The language "phosphatidylinositol activity", as used herein, refers to an activity of $PIP_2$ or one of its metabolites.

**[0054]** Another signaling pathway in which the receptor may participate is the cAMP turnover pathway. As used herein, "cyclic AMP turnover and metabolism" refers to the molecules involved in the turnover and metabolism of cyclic AMP (cAMP) as well as to the activities of these molecules. Cyclic AMP is a second messenger produced in response to ligand-induced stimulation of certain G protein coupled receptors. In the cAMP signaling pathway, binding of a ligand to a GPCR can lead to the activation of the enzyme adenyl cyclase, which catalyzes the synthesis of cAMP. The newly synthesized cAMP can in turn activate a cAMP-dependent protein kinase. This activated kinase can phosphorylate a voltage-gated potassium channel protein, or an associated protein, and lead to the inability of the potassium channel to open during an action potential. The inability of the potassium channel to open results in a decrease in the outward flow of potassium, which normally repolarizes the membrane of a neuron, leading to prolonged membrane depolarization.



10085232.022802

**[0055]** The 93870 molecules and modulators thereof can act as novel therapeutic agents for controlling one or more of GPCR associated disorders, e.g., immune and inflammatory disorders, platelet disorders, skeletal or bone metabolism disorders, and bone marrow mononuclear cell disorders, as described herein.

**[0056]** GPCR-associated disorders can include immune and inflammatory diseases (including, for example, diabetes mellitus, arthritis (including rheumatoid arthritis, juvenile rheumatoid arthritis, osteoarthritis, psoriatic arthritis), multiple sclerosis, encephalomyelitis, myasthenia gravis, systemic lupus erythematosis, autoimmune thyroiditis, dermatitis (including atopic dermatitis and eczematous dermatitis), psoriasis, Sjögren's Syndrome, inflammatory bowel disease (e.g., Crohn's disease and ulcerative colitis), aphthous ulcer, iritis, conjunctivitis, keratoconjunctivitis, respiratory inflammation (e.g., asthma, allergic asthma, and chronic obstructive pulmonary disease), cutaneous lupus erythematosus, scleroderma, vaginitis, proctitis, drug eruptions, leprosy reversal reactions, erythema nodosum leprosum, autoimmune uveitis, allergic encephalomyelitis, acute necrotizing hemorrhagic encephalopathy, idiopathic bilateral progressive sensorineural hearing loss, aplastic anemia, pure red cell anemia, idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis, chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue, lichen planus, Graves' disease, sarcoidosis, primary biliary cirrhosis, uveitis posterior, and interstitial lung fibrosis), graft-versus-host disease, cases of transplantation, respiratory inflammation(consisting of asthma and chronic obstructive pulmonary diseas) and allergy such as, atopic allergy. Other examples of GPCR associated immune and inflammatory disorders are congenital X-linked infantile hypogammaglobulinemia, transient hypogammaglobulinemia, common variable immunodeficiency, selective IgA deficiency, chronic mucocutaneous candidiasis, or severe combined immunodeficiency.

**[0057]** Blood platelet disorders include, but are not limited to, thrombocytopenia due to a reduced number of megakaryocytes in the bone marrow, for example, as a result of chemotherapy; invasive disorders, such as leukemia, idiopathic or drug- or toxin-induced aplasia of the marrow, or rare hereditary amegakaryocytic thrombocytopenias; ineffective thrombopoiesis, for example, as a result of megaloblastic anemia, alcohol toxicity, vitamin B12 or folate deficiency, myelodysplastic disorders, or rare hereditary disorders (e.g., Wiskott-Aldrich syndrome and May-hegglin anomaly); a reduction in platelet distribution, for example, as a result of cirrhosis, a splenic invasive disease (e.g., Gaucher's disease), or myelofibrosis with extramedullary myeloid metaplasia; increased platelet destruction, for

example, as a result of removal of IgG-coated platelets by the mononuclear phagocytic system (e.g., idiopathic thrombocytopenic purpura (ITP), secondary immune thrombocytopenia (e.g., systemic lupus erythematosus, lymphoma, or chronic lymphocytic leukemia), drug-related immune thrombocytopenias (e.g., as with quinidine, aspirin, and heparin), post-transfusion purpura, and neonatal thrombocytopenia as a result of maternal platelet autoantibodies or maternal platelet alloantibodies). Also included are thrombocytopenia secondary to intravascular clotting and thrombin induced damage to platelets as a result of, for example, obstetric complications, metastatic tumors, severe gram-negative bacteremia, thrombotic thrombocytopenic purpura, or severe illness. Also included is dilutional thrombocytopenia, for example, due to massive hemorrhage. Blood platelet disorders also include, but are not limited to, essential thrombocytosis and thrombocytosis associated with, for example, splenectomy, acute or chronic inflammatory diseases, hemolytic anemia, carcinoma, Hodgkin's disease, lymphoproliferative disorders, and malignant lymphomas.

**[0058]** Aberrant expression and/or activity of 93870 molecules can mediate disorders associated with skeletal integrity or bone metabolism. “Skeletal integrity” refers to direct or indirect effects on the formation or maintenance of bones or joints and the tissues, such as ligaments, tendons or muscles which connect and control movement of those structures. “Bone metabolism” refers to direct or indirect effects in the formation or degeneration of bone structures, e.g., bone formation, bone resorption, etc., which may ultimately affect the structural integrity of bones or the concentrations in serum of calcium and phosphate. This term also includes activities mediated by 93870 molecule effects in bone cells, e.g. osteoclasts and osteoblasts, that may in turn result in bone formation and degeneration. These cells are responsible for the synthesis and remodeling of the collagenous bone matrix, among other activities. Collagen is the main structural component of bone and bone-associated skeletal structures beside calcium phosphate, so collagen synthesis and processing are integral to bone or skeletal structure. 93870 molecules can be involved in one or more of the steps which result in the correct collagen matrix for bone. In another example, 93870 molecules can support different activities of bone resorbing osteoclasts such as the stimulation of differentiation of monocytes and mononuclear phagocytes into osteoclasts. Accordingly, 93870 molecules that modulate the production of bone cells can influence bone formation and degeneration, and thus can be used to treat skeletal or bone disorders. Examples of such disorders include, but are not limited to, osteogenesis

10085233.022802

imperfecta, osteoporosis, osteodystrophy, osteomalacia, rickets, osteitis fibrosa cystica, Ehlers-Danlos syndrome, renal osteodystrophy, osteosclerosis, anti- convulsant treatment, osteopenia, fibrogenesis-imperfecta ossium, secondary hyperparathyrodism, hypoparathyroidism, hyperparathyroidism, cirrhosis, obstructive jaundice, drug induced metabolism, medullary carcinoma, chronic renal disease, rickets, sarcoidosis, glucocorticoid antagonism, malabsorption syndrome, steatorrhea, tropical sprue, idiopathic hypercalcemia and milk fever.

**[0059]** The 93870 protein, fragments thereof, and derivatives and other variants of the sequence in SEQ ID NO:2 thereof are collectively referred to as "polypeptides or proteins of the invention" or "93870 polypeptides or proteins". Nucleic acid molecules encoding such polypeptides or proteins are collectively referred to as "nucleic acids of the invention" or “93870 nucleic acids."

**[0060]** As used herein, the term "nucleic acid molecule" includes DNA molecules (e.g., a cDNA or genomic DNA) and RNA molecules (e.g., an mRNA) and analogs of the DNA or RNA generated, e.g., by the use of nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

**[0061]** The term "isolated or purified nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of 5’ and/or 3’ nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

**[0062]** As used herein, the term "hybridizes under stringent conditions" describes conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons,

N.Y. (1989), 6.3.1-6.3.6. Aqueous and nonaqueous methods are described in that reference and either can be used. A preferred, example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Preferably, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. Particularly preferred stringency conditions (and the conditions that should be used if the practitioner is uncertain about what conditions should be applied to determine if a molecule is within a hybridization limitation of the invention) are 0.5M Sodium Phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1 or SEQ ID NO:3, corresponds to a naturally-occurring nucleic acid molecule.

**[0063]** As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

**[0064]** As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a 93870 protein, preferably a mammalian 93870 protein, and can further include non-coding regulatory sequences, and introns.

**[0065]** An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language "substantially free" means preparation of 93870 protein having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-93870 protein (also referred to herein as a "contaminating protein"), or of chemical precursors or non-93870 chemicals. When the 93870 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%,

10085233.082202

more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The invention includes isolated or purified preparations of at least 0.01, 0.1, 1.0, and 10 milligrams in dry weight.

**[0066]** A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of 93870 (e.g., the sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____) without abolishing or more preferably, without substantially altering a biological activity, whereas an "essential" amino acid residue results in such a change. For example, amino acid residues that are conserved among the polypeptides of the present invention, e.g., those present in the seven transmembrane receptor domains, are predicted to be particularly unamenable to alteration.

**[0067]** A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a 93870 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a 93870 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for 93870 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1 or SEQ ID NO:3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

**[0068]** As used herein, a "biologically active portion" of a 93870 protein includes a fragment of a 93870 protein which participates in an interaction between a 93870 molecule and a non-93870 molecule. Biologically active portions of a 93870 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the 93870 protein, e.g., the amino acid sequence shown in SEQ ID NO:2, which

include less amino acids than the full length 93870 proteins, and exhibit at least one activity of a 93870 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the 93870 protein, e.g., a domain or motif capable of regulating, sensing and/or transmitting an extracellular signal into a cell, for example, a hematopoietic cell; a domain or motif capable of interacting with (e.g., binding to) an extracellular signal or a cell surface receptor; a domain or motif capable of mobilizing an intracellular molecule that participates in a signal transduction pathway (e.g., adenylate cyclase or phosphatidylinositol 4,5-bisphosphate ($PIP_2$), inositol 1,4,5-triphosphate ($IP_3$)); a domain or motif capable of regulating polarization of the plasma membrane; a domain or motif capable of controlling production or secretion of molecules; a domain or motif capable of altering the structure of a cellular component; a domain or motif capable of modulating cell proliferation, e.g., synthesis of DNA; and/or a domain or motif capable of modulating migration, proliferation and/or differentiation of a cell.

**[0069]** A biologically active portion of a 93870 protein can be a polypeptide which is, for example, 10, 25, 50, 100, 200 or more amino acids in length. Biologically active portions of a 93870 protein can be used as targets for developing agents which modulate a 93870 mediated activity, e.g., a biological activity described herein.

**[0070]** Calculations of homology or sequence identity (the terms "homology and identity" are used interchangeably herein) between sequences are performed as follows.

**[0071]** To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence (e.g., when aligning a second sequence to the 93870 amino acid sequence of SEQ ID NO:2 having 313 amino acid residues, at least 94, preferably at least 125, more preferably at least 157, even more preferably at least 189, and even more preferably at least 219, 250, 282, or 313 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then

the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

**[0072]** The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be applied to determine if a molecule is within a sequence identity or homology limitation of the invention) are a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

10085233.022802

**[0073]** The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

**[0074]** The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 93870 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to 93870 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) *Nucleic Acids Res.* 25(17):3389-3402.

When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

**[0075]** "Misexpression or aberrant expression", as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein level. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-transitional modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

**[0076]** "Subject", as used herein, can refer to a mammal, e.g., a human, or to an experimental or animal or disease model. The subject can also be a non-human animal, e.g., a horse, cow, goat, or other domestic animal.

**[0077]** A "purified preparation of cells", as used herein, refers to, in the case of plant or animal cells, an in vitro preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

**[0078]** Various aspects of the invention are described in further detail below.

Isolated Nucleic Acid Molecules

**[0079]** In one aspect, the invention provides an isolated or purified nucleic acid molecule that encodes a 93870 polypeptide described herein, e.g., a full-length 93870 protein or a fragment thereof, e.g., a biologically active portion of 93870 protein. Also included is a nucleic acid fragment suitable for use as a hybridization probe, which can be used, e.g., to a identify nucleic acid molecule encoding a polypeptide of the invention, 93870 mRNA, and fragments suitable for use as primers, e.g., PCR primers for the amplification or mutation of nucleic acid molecules.

10085233.022802

**[0080]** In one embodiment, an isolated nucleic acid molecule of the invention includes the nucleotide sequence shown in SEQ ID NO:1, or the deposited nucleotide sequence, or a portion of either of these nucleotide sequences. In one embodiment, the nucleic acid molecule includes sequences encoding the human 93870 protein (i.e., "the coding region," from nucleotides 147-1085 of SEQ ID NO:1), as well as 5'-untranslated sequences (nucleotides 1-146 of SEQ ID NO:1) or 3'-untranslated sequences (nucleotides 1086-1684 of SEQ ID NO:1). Alternatively, the nucleic acid molecule can include only the coding region of SEQ ID NO:1 (e.g., nucleotides 147-1085 of SEQ ID NO:1, corresponding to nucleotides 1-939 SEQ ID NO:3) and, e.g., no flanking sequences which normally accompany the subject sequence. In another embodiment, the nucleic acid molecule encodes a sequence corresponding to the 313 amino acid residue protein of SEQ ID NO:2.

**[0081]** In another embodiment, an isolated nucleic acid molecule of the invention includes a nucleic acid molecule which is a complement of the nucleotide sequence shown in one of SEQ ID NO:1, SEQ ID NO:3, the deposited nucleotide sequence, or a portion of any of these sequences. In other embodiments, the nucleic acid molecule of the invention is sufficiently complementary to the nucleotide sequence shown in one of SEQ ID NO:1, SEQ ID NO:3, and the deposited nucleotide sequence that it can hybridize with a nucleic acid having that sequence, thereby forming a stable duplex.

**[0082]** In one embodiment, an isolated nucleic acid molecule of the invention includes a nucleotide sequence which is at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more homologous to the entire length of the nucleotide sequence shown in one of SEQ ID NO:1, SEQ ID NO:3, the deposited nucleotide sequence, and a portion, preferably of the same length, of any of these nucleotide sequences.

<u>93870 Nucleic Acid Fragments</u>

**[0083]** A nucleic acid molecule of the invention can include only a portion of the nucleic acid sequence of one of SEQ ID NOs:1 and 3 and the deposited nucleotide sequence. For example, such a nucleic acid molecule can include a fragment that can be used as a probe or primer or a fragment encoding a portion of a 93870 protein, e.g., an immunogenic or biologically active portion of a 93870 protein. A fragment can comprise nucleotides encoding a fragment corresponding to residues 56-123 of SEQ ID NO:2, which is a receptor coupled G-protein transmembrane domain of human 93870. The nucleotide sequence

determined from the cloning of the 93870 gene facilitates generation of probes and primers for use in identifying and/or cloning other 93870 family members, or fragments thereof, as well as 93870 homologues, or fragments thereof, from other species.

**[0084]** In another embodiment, a nucleic acid includes a nucleotide sequence that includes part, or all, of the coding region and extends into either (or both) the 5'- or 3'-non-coding region. Other embodiments include a fragment that includes a nucleotide sequence encoding an amino acid fragment described herein. Nucleic acid fragments can encode a specific domain or site described herein or fragments thereof, particularly fragments thereof that are at least about 263, 265, 270, 275, 280, or more amino acids in length. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments should not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

10085233.022802

**[0085]** A nucleic acid fragment can include a sequence corresponding to a domain, region, or functional site described herein. A nucleic acid fragment can also include one or more domain, region, or functional site described herein.

**[0086]** 93870 probes and primers are provided. Typically a probe/primer is an isolated or purified oligonucleotide. The oligonucleotide typically includes a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense or antisense sequence of one of SEQ ID NO:1, SEQ ID NO:3, the deposited nucleotide sequence, and a naturally occurring allelic variant or mutant of SEQ ID NO:1, SEQ ID NO:3, or the deposited nucleotide sequence.

**[0087]** In a preferred embodiment the nucleic acid is a probe which is at least 5 or 10, and less than 200, more preferably less than 100, or less than 50, base pairs in length. It should be identical, or differ by 1, or fewer than 5 or 10 bases, from a sequence disclosed herein. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

**[0088]** A probe or primer can be derived from the sense or anti-sense strand of a nucleic acid that encodes, for example, an receptor coupled G-protein transmembrane domain at about amino acid residues 56 to 123 of SEQ ID NO:2

**[0089]** In another embodiment a set of primers is provided, e.g., primers suitable for use in a PCR, which can be used to amplify a selected region of a 93870 sequence. The primers

should be at least 5, 10, or 50 base pairs in length and less than 100- 200, base pairs in length. The primers should be identical, or differs by one base from a sequence disclosed herein or from a naturally occurring variant. Primers suitable for amplifying all or a portion of any of the following regions of a receptor coupled G-protein transmembrane domain are provided: for example, any of the transmembrane domains of the 93870 polypeptide (i.e., amino acid residues 28-51, 58-79, 98-119, 140-160, 192-216, 236-252, and 276-299 of SEQ ID NO:2), or any of the extracellular domains of the 93870 polypeptide (i.e., amino acid residues 80-97, 161-191, and 253-275) of SEQ ID NO:2, or any of the the intracellular domains of the 93870 polypeptide (i.e., amino acid residues 52-59, 120-139, and 217-235) of SEQ ID NO:2.

**[0090]** A nucleic acid fragment can encode an epitope bearing region of a polypeptide described herein.

**[0091]** A nucleic acid fragment encoding a "biologically active portion of a 93870 polypeptide" can be prepared by isolating a portion of the nucleotide sequence of one of SEQ ID NO:1, SEQ ID NO:3, and the deposited nucleotide sequence, which encodes a polypeptide having a 93870 biological activity (e.g., the biological activities of the 93870 proteins as described herein), expressing the encoded portion of the 93870 protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the 93870 protein. For example, a nucleic acid fragment encoding a biologically active portion of 93870 can include a receptor coupled G-protein transmembrane domain, e.g., amino acid residues 56 to 123 of SEQ ID NO:2. A nucleic acid fragment encoding a biologically active portion of a 93870 polypeptide can comprise a nucleotide sequence that is greater than 25 or more nucleotides in length.

**[0092]** In one embodiment, a nucleic acid includes a nucleotide sequence which is greater than 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 750, 800, 850, 900, or more nucleotides in length and that hybridizes under stringent hybridization conditions with a nucleic acid molecule having the sequence of one of SEQ ID NO:1, SEQ ID NO:3, and the deposited nucleotide sequence.

<u>93870 Nucleic Acid Variants</u>

**[0093]** The invention further encompasses nucleic acid molecules having a sequence that differs from the nucleotide sequence shown in one of SEQ ID NO:1, SEQ ID NO:3, and the deposited nucleotide sequence. Such differences can be attributable to degeneracy of the



10085253.022802

genetic code (i.e., differences which result in a nucleic acid that encodes the same 93870 proteins as those encoded by the nucleotide sequence disclosed herein). In another embodiment, an isolated nucleic acid molecule of the invention encodes a protein having an amino acid sequence which differs by at least 1, but by fewer than 5, 10, 20, 50, or 100, amino acid residues from SEQ ID NO:2. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

**[0094]** Nucleic acids of the invention can be chosen for having codons, which are preferred, or non-preferred, for a particular expression system. For example, the nucleic acid can be one in which at least one codon, preferably at least 10%, or 20% of the codons has been altered such that the sequence is optimized for expression in *E. coli*, yeast, human, insect, or CHO cells.

**[0095]** Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism), or can be non-naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product).

**[0096]** In a preferred embodiment, the nucleic acid has a sequence that differs from that of one of SEQ ID NO:1, SEQ ID NO:3, and the deposited nucleotide sequence, e.g., as follows: by at least one, but fewer than 10, 20, 30, or 40, nucleotide residues; or by at least one but fewer than 1%, 5%, 10% or 20% of the nucleotide residues in the subject nucleic acid. If necessary for this analysis, the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

**[0097]** Orthologs, homologs, and allelic variants can be identified using methods known in the art. These variants comprise a nucleotide sequence encoding a polypeptide that is 50%, at least about 55%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more identical to the nucleotide sequence shown in one of SEQ ID NO:1, SEQ ID NO:3, the deposited nucleotide sequence, or a fragment of one of these sequences. Such nucleic acid molecules can readily be identified as

10085233.022802

being able to hybridize under stringent conditions to the nucleotide sequence of one of SEQ ID NO:1, SEQ ID NO:3, the deposited nucleotide sequence, or a fragment of one of these sequences. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of the 93870 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the 93870 gene.

**[0098]** Preferred variants include those that are correlated with any of the 93870 biological activities described herein, e.g., transducing a chemical or protein signal via a G-protein coupled receptor.

**[0099]** Allelic variants of 93870 (e.g., human 93870) include both functional and non-functional proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the 93870 protein within a population that maintain the ability to mediate any of the 93870 biological activities described herein.

**[00100]** Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:2, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein. Non-functional allelic variants are naturally-occurring amino acid sequence variants of the 93870 (e.g., human 93870) protein within a population that do not have the ability to mediate any of the 93870 biological activities described herein. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion, or premature truncation of the amino acid sequence of SEQ ID NO:2, or a substitution, insertion, or deletion in critical residues or critical regions of the protein.

**[00101]** Moreover, nucleic acid molecules encoding other 93870 family members and, thus, which have a nucleotide sequence which differs from the 93870 sequences of one of SEQ ID NO:1, SEQ ID NO:3, and the deposited nucleotide sequence are within the scope of the invention.

<u>Antisense Nucleic Acid Molecules, Ribozymes and Modified 93870 Nucleic Acid Molecules</u>

**[00102]** In another aspect, the invention features, an isolated nucleic acid molecule that is antisense to 93870. An "antisense" nucleic acid can include a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire 93870 coding

strand, or to only a portion thereof (e.g., the coding region of human 93870 corresponding to SEQ ID NO:3). In another embodiment, the antisense nucleic acid molecule is antisense to a "non-coding region" of the coding strand of a nucleotide sequence encoding 93870 (e.g., the 5'- and 3'-untranslated regions).

**[00103]** An antisense nucleic acid can be designed such that it is complementary to the entire coding region of 93870 mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or non-coding region of 93870 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of 93870 mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, or 80 or more nucleotide residues in length.

**[00104]** An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. The antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been sub-cloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

**[00105]** The antisense nucleic acid molecules of the invention are typically administered to a subject (e.g., by direct injection at a tissue site), or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 93870 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular



10085233.022802

concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

**[00106]** In yet another embodiment, the antisense nucleic acid molecule of the invention is an alpha-anomeric nucleic acid molecule. An alpha-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual beta-units, the strands run parallel to each other (Gaultier *et al.* (1987) Nucl. *Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucl. Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS* Lett. 215:327-330).

**[00107]** In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. A ribozyme having specificity for a 93870-encoding nucleic acid can include one or more sequences complementary to the nucleotide sequence of a 93870 cDNA disclosed herein (i.e., SEQ ID NO:1 or SEQ ID NO:3), and a sequence having known catalytic sequence responsible for mRNA cleavage (see, for example, U.S. Patent number 5,093,246 or Haselhoff *et al.* (1988, *Nature* 334:585-591). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a 93870-encoding mRNA (e.g., U.S. Patent number 4,987,071; and U.S. Patent number 5,116,742). Alternatively, 93870 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (e.g., Bartel *et al.*(1993) *Science* 261:1411-1418).

**[00108]** 93870 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 93870 (e.g., the 93870 promoter and/or enhancers) to form triple helical structures that prevent transcription of the 93870 gene in target cells (Helene, 1991, Anticancer Drug Des. 6:569-584; Helene, *et al.*(1992) *Ann. N.Y. Acad. Sci.* 660:27-36; Maher(1992) *Bioassays* 14:807-815). The potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5' to 3', 3' to 5' manner, such that they hybridize with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

**[00109]** The invention also provides detectably labeled oligonucleotide primer and probe molecules. Typically, such labels are chemiluminescent, fluorescent, radioactive, or colorimetric.

**[00110]** A 93870 nucleic acid molecule can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (Hyrup *et al.*(1996*) Bioorg. Med. Chem.* 4:5-23). As used herein, the terms "peptide nucleic acid" (PNA) refers to a nucleic acid mimic, e.g., a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) *supra*; Perry-O'Keefe *et al., Proc. Natl. Acad. Sci. USA* 93:14670-14675.

**[00111]** PNAs of 93870 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or anti-gene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of 93870 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (e.g., S1 nucleases, as described in Hyrup *et al.* (1996) *supra*; or as probes or primers for DNA sequencing or hybridization (Hyrup *et al*(1996) *supra*; Perry-O'Keefe, *supra*).

**[00112]** In other embodiments, the oligonucleotide can include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (e.g., Letsinger *et al* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.*, (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT publication number WO 88/09810) or the blood-brain barrier (see, e.g., PCT publication number WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (e.g., Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents (e.g., Zon, 1988, Pharm. Res. 5:539-549). To this end, the oligonucleotide can be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

**[00113]** The invention also includes molecular beacon oligonucleotide primer and probe molecules having at least one region which is complementary to a 93870 nucleic acid of the invention, two complementary regions, one having a fluorophore and the other having a quencher, such that the molecular beacon is useful for quantitating the presence of the 93870

10085233.022802

nucleic acid of the invention in a sample. Molecular beacon nucleic acids are described, for example, in U.S. Patent number. 5,854,033, U.S. Patent number 5,866,336, and U.S. Patent number 5,876,930.

Isolated 93870 Polypeptides

**[00114]** In another aspect, the invention features, an isolated 93870 protein, or fragment, e.g., a biologically active portion, for use as immunogens or antigens to raise or test (or more generally to bind) anti-93870 antibodies. 93870 protein can be isolated from cells or tissue sources using standard protein purification techniques. 93870 protein or fragments thereof can be produced by recombinant DNA techniques or synthesized chemically.

**[00115]** Polypeptides of the invention include those that arise as a result of the existence of multiple genes, alternative transcription events, alternative RNA splicing events, and alternative translational and post-translational events. The polypeptide can be expressed in systems, e.g., cultured cells, which result in substantially the same post-translational modifications present when the polypeptide is expressed in a native cell, or in systems which result in the alteration or omission of post-translational modifications, e.g., glycosylation or cleavage, present when expressed in a native cell.

**[00116]** In a preferred embodiment, a 93870 polypeptide has one or more of the following characteristics described in the art (e.g., Strader *et al., supra*, and references cited therein). A preferred embodiment of a 93870 polypeptide also has one or more of the following characteristics:

**[00117]** it has the ability to regulate, sense and/or transmit an extracellular signal into a cell;

**[00118]** it has the ability to interact with (e.g., bind to) an extracellular signal or a cell surface receptor;

**[00119]** it has the ability to mobilize an intracellular molecule that participates in a signal transduction pathway (e.g., adenylate cyclase or phosphatidylinositol 4,5-bisphosphate ($PIP_2$), inositol 1,4,5-triphosphate ($IP_3$));

**[00120]** it has the ability to modulate proliferation, migration, differentiation and/or survival of a cell;

**[00121]** it has the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which it is expressed e.g, bone marrow mononuclear cells, neutrophils, osteoblasts, and megakaryocytes;

**[00122]** it has a molecular weight, amino acid composition or other physical characteristic of a 93870 protein of SEQ ID NO:2;

**[00123]** it has an overall sequence similarity (identity) of at least 60-65%, preferably at least 70%, more preferably at least 75, 80, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% or more, with a polypeptide of SEQ ID NO:2;

**[00124]** it has an N-terminal extracellular domain which is preferably about 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or more, identical with amino acid residues 1-27 of SEQ ID NO:2;

**[00125]** it has at least one transmembrane domains which is preferably about 70%, 80%, 90%, 95% or higher, identical with amino acid residues 28-51, 58-79, 98-119, 140-160, 192-216, 236-252, 276-299 of SEQ ID NO:2;

**[00126]** it has a C-terminal domain which is preferably about 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or more, identical with amino acid residues 300-313 of SEQ ID NO:2; or

**[00127]** it has a motif identical to the arginine-aromatic motif present at amino acids residues about 121-122 of SEQ ID NO:2.

**[00128]** In a preferred embodiment, the 93870 protein or fragment thereof differs only insubstantially, if at all, from the corresponding sequence in SEQ ID NO:2. In one embodiment, it differs by at least one, but by fewer than 15, 10 or preferably 5 amino acid residues. In another, it differs from the corresponding sequence in SEQ ID NO:2 by at least one residue but fewer than 20%, 15%, 10% or 5% of the residues differ from the corresponding sequence in SEQ ID NO:2 (if this comparison requires alignment the sequences should be aligned for maximum homology). "Looped" out sequences from deletions or insertions, or mismatches, are considered differences). The differences are, preferably, differences or changes at a non-essential amino acid residues or involve a conservative substitution of one residue for another. In a preferred embodiment the differences are not in a transmembrane domain, the extracellular domain, or the second cytoplasmic loop of the 93870 protein N-terminal.

**[00129]** Other embodiments include a protein that has one or more changes in amino acid sequence, relative to SEQ ID NO:2 (e.g., a change in an amino acid residue which is not essential for activity). Such 93870 proteins differ in amino acid sequence from SEQ ID NO:2, yet retain biological activity.



10085233.022802

**[00130]** An expression pattern as follows, based on TaqMan analysis of 93870 expression in at least the following human tissues and cell lines (described below in the Exemplification): high levels of 93870 expression in normal bone marrow mononuclear cells, and neutrophils; lower levels of 93870 expression in human osteoblasts, and megakaryocytes; and

**[00131]** In one embodiment, the protein includes an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more homologous to SEQ ID NO:2.

**[00132]** A 93870 protein or fragment is provided which has an amino acid sequence which varies from SEQ ID NO:2 in one or both of the regions corresponding to, for example residues 120-139 of SEQ ID NO:2 by at least one, but by fewer than 15, 10 or 5 amino acid residues, but which does not differ from SEQ ID NO:2 in the regions corresponding to the transmembrane domain, the N-terminal extracellular domain, or the second cytoplasmic loop of the 93870 protein (if this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences). In some embodiments the difference is at a non-essential residue or is a conservative substitution, while in others the difference is at an essential residue or is a non-conservative substitution.

**[00133]** A preferred biologically active portion of a 93870 protein can include at least one or more of the following: An N-terminal extracellular domain, an extracellular loop, and a cytoplasmic loop. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques (e.g., to create chimeric receptors by replacing an intracellular domain of the 93870 protein with that of a characterized domain of a protein known in the art that initiates a defined signal transduction cascade) and evaluated for one or more of the functional activities of a native 93870 protein.

**[00134]** In a preferred embodiment, the 93870 protein has the amino acid sequence SEQ ID NO:2. In other embodiments, the 93870 protein is substantially identical to SEQ ID NO:2. In yet another embodiment, the 93870 protein is substantially identical to SEQ ID NO:2 and retains the functional activity of the protein of SEQ ID NO:2.

<u>93870 Chimeric or Fusion Proteins</u>

**[00135]** In another aspect, the invention provides 93870 chimeric or fusion proteins. As used herein, a 93870 "chimeric protein" or "fusion protein" includes a 93870 polypeptide linked to a non-93870 polypeptide. A "non-93870 polypeptide" refers to a polypeptide



10085233.022802

having an amino acid sequence corresponding to a protein which is not substantially homologous to the 93870 protein, e.g., a protein which is different from the 93870 protein and which is derived from the same or a different organism. The 93870 polypeptide of the fusion protein can correspond to all or a portion e.g., a fragment described herein of a 93870 amino acid sequence. In a preferred embodiment, a 93870 fusion protein includes at least one or more biologically active portions of a 93870 protein. The non-93870 polypeptide can be fused to the amino or carboxyl terminus of the 93870 polypeptide.

**[00136]** The fusion protein can include a moiety that has a high affinity for a ligand. For example, the fusion protein can be a GST-93870 fusion protein in which the 93870 sequences are fused to the carboxyl terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant 93870. Alternatively, the fusion protein can be a 93870 protein containing a heterologous signal sequence at its amino terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of 93870 can be increased through use of a heterologous signal sequence.

**[00137]** Fusion proteins can include all or a part of a serum protein, e.g., a portion of an immunoglobulin (e.g., IgG, IgA, or IgE), e.g., an Fc region and/or the hinge C1 and C2 sequences of an immunoglobulin or human serum albumin.

**[00138]** The 93870 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The 93870 fusion proteins can be used to affect the bioavailability of a 93870 substrate. 93870 fusion proteins can be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a 93870 protein; (ii) mis-regulation of the 93870 gene; and (iii) aberrant post-translational modification of a 93870 protein.

**[00139]** Moreover, the 93870-fusion proteins of the invention can be used as immunogens to produce anti-93870 antibodies in a subject, to purify 93870 ligands and in screening assays to identify molecules that inhibit the interaction of 93870 with a 93870 substrate.

**[00140]** Expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A 93870-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 93870 protein.

Variants of 93870 Proteins

**[00141]** In another aspect, the invention also features a variant of a 93870 polypeptide, e.g., which functions as an agonist (mimetics) or as an antagonist. Variants of the 93870 proteins can be generated by mutagenesis, e.g., discrete point mutation, the insertion or deletion of sequences or the truncation of a 93870 protein. An agonist of the 93870 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a 93870 protein. An antagonist of a 93870 protein can inhibit one or more of the activities of the naturally occurring form of the 93870 protein by, for example, competitively modulating a 93870-mediated activity of a 93870 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Preferably, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the 93870 protein.

**[00142]** Variants of a 93870 protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of a 93870 protein for agonist or antagonist activity.

**[00143]** Libraries of fragments e.g., amino-terminal, carboxyl-terminal, or internal fragments, of a 93870 protein coding sequence can be used to generate a variegated population of fragments for screening and subsequent selection of variants of a 93870 protein.

**[00144]** Variants in which a cysteine residue is added or deleted or in which a residue that is glycosylated is added or deleted are particularly preferred.

**[00145]** Methods for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify 93870 variants (Arkin *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engr.* 6:327-331).

**[00146]** Cell based assays can be exploited to analyze a variegated 93870 library. For example, a library of expression vectors can be transfected into a cell line, e.g., a cell line, which ordinarily responds to 93870 in a substrate-dependent manner. The transfected cells are then contacted with 93870 and the effect of the expression of the mutant on signaling by the 93870 substrate can be detected, e.g., by measuring changes in cell growth and/or enzymatic activity. Plasmid DNA can then be recovered from the cells that score for



10085233.022802

inhibition, or alternatively, potentiation of signaling by the 93870 substrate, and the individual clones further characterized.

**[00147]** In another aspect, the invention features a method of making a 93870 polypeptide, e.g., a peptide having a non-wild-type activity, e.g., an antagonist, agonist, or super agonist of a naturally-occurring 93870 polypeptide, e.g., a naturally-occurring 93870 polypeptide. The method includes: altering the sequence of a 93870 polypeptide, e.g., altering the sequence, e.g., by substitution or deletion of one or more residues of a non-conserved region, a domain or residue disclosed herein, and testing the altered polypeptide for the desired activity.

**[00148]** In another aspect, the invention features a method of making a fragment or analog of a 93870 polypeptide a biological activity of a naturally occurring 93870 polypeptide. The method includes: altering the sequence, e.g., by substitution or deletion of one or more residues, of a 93870 polypeptide, e.g., altering the sequence of a non-conserved region, or a domain or residue described herein, and testing the altered polypeptide for the desired activity.

<u>Anti-93870 Antibodies</u>

**[00149]** In another aspect, the invention provides an anti-93870 antibody. The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include scFV and dcFV fragments, F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as papain or pepsin respectively.

**[00150]** The antibody can be a polyclonal, monoclonal, recombinant, e.g., a chimeric, humanized, fully-human, non-human, e.g., murine, or single chain antibody. In a preferred embodiment, it has effector function and can fix complement. The antibody can be coupled to a toxin or imaging agent.

**[00151]** A full-length 93870 protein or, antigenic peptide fragment of 93870 can be used as an immunogen or can be used to identify anti-93870 antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. The antigenic peptide of 93870 should include at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of 93870. Preferably, the antigenic peptide includes at least 10 amino acid residues, more preferably at least 15 amino acid residues,

even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

**[00152]** Fragments of 93870 which include an extracellular domain or region or a cytoplasmic domain or region as described herein, of SEQ ID NO:2 can be used to make antibodies. Alternatively, a fragment of 93870 which includes a transmembrane domain of SEQ ID NO:2, as described herein, can be used to make an antibody. Figure 1 depicts a hydropathy plot of human 93870 which can be used to show areas of hydrophobicity or hydrophilicity, against which other anti-93870 antibodies can be made.

**[00153]** Antibodies reactive with, or specific or selective for, any of these regions, or other regions or domains described herein are provided.

**[00154]** Preferred epitopes encompassed by the antigenic peptide are regions of 93870 are located on the surface of the protein, e.g., hydrophilic regions (see Figure 1), as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human 93870 protein sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the 93870 protein and are thus likely to constitute surface residues useful for targeting antibody production. In a preferred embodiment the antibody binds an epitope on any domain or region on 93870 proteins described herein.

**[00155]** In a preferred embodiment the antibody binds an epitope on any domain or region on 93870 proteins described herein.

**[00156]** Chimeric, humanized, but most preferably, completely human antibodies are desirable for applications which include repeated administration, e.g., therapeutic treatment (and some diagnostic applications) of human patients.

**[00157]** The anti-93870 antibody can be a single chain antibody. A single-chain antibody (scFV) can be engineered as described in for example, Colcher *et al.* (1999) *Ann. N.Y. Acad. Sci.* 880:263-280; and Reiter (1996) *Clin. Cancer Res.* 2:245-252. The single chain antibody can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target 93870 protein.

**[00158]** In a preferred embodiment, the antibody has reduced or no ability to bind an Fc receptor. For example, it can be an isotype, subtype, fragment or other mutant, which does not support binding to an Fc receptor, e.g., it can have a mutated or deleted Fc receptor binding region.

**[00159]** An anti-93870 antibody (e.g., monoclonal antibody) can be used to isolate 93870 by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an anti-93870 antibody can be used to detect 93870 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the protein. Anti-93870 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, beta-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include $^{125}I$, $^{131}I$, $^{35}S$ or $^{3}H$.

10085233.022802

<u>Recombinant Expression Vectors, Host Cells and Genetically Engineered Cells</u>

**[00160]** In another aspect, the invention includes, vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide described herein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors include, e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses.

**[00161]** A vector can include a 93870 nucleic acid in a form suitable for expression of the nucleic acid in a host cell. Preferably the recombinant expression vector includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the

like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or polypeptides, including fusion proteins or polypeptides, encoded by nucleic acids as described herein (e.g., 93870 proteins, mutant forms of 93870 proteins, fusion proteins, and the like).

**[00162]** The recombinant expression vectors of the invention can be designed for expression of 93870 proteins in prokaryotic or eukaryotic cells. For example, polypeptides of the invention can be expressed in *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel (1990, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

**[00163]** Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith *et al.* (1988) Gene 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

**[00164]** Purified fusion proteins can be used in 93870 activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific for 93870 proteins. In a preferred embodiment, a fusion protein expressed in a retroviral expression vector of the present invention can be used to infect bone marrow cells that are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six weeks).

**[00165]** To maximize recombinant protein expression in *E. coli*, the protein is expressed in a host bacterial strain with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, 1990, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.* (1992) *Nucl. Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

**[00166]** The 93870 expression vector can be a yeast expression vector, a vector for expression in insect cells, e.g., a baculovirus expression vector, or a vector suitable for expression in mammalian cells.

**[00167]** When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used viral promoters are derived from polyoma, adenovirus 2, cytomegalovirus and simian virus 40 (SV40).

**[00168]** In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame *et al.* (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto *et al.* (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) Cell 33:729-740; Queen *et al.* (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent number 4,873,316 and European Patent Application publication number 264,166). Developmentally-regulated promoters are also encompassed, for example, the murine hox promoters (Kessel *et al.* (1990) *Science* 249:374-379) and the alpha-fetoprotein promoter (Campes *et al.* (1989) *Genes Dev.* 3:537-546).

**[00169]** The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. The

10085233.022802

antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the regulation of gene expression using antisense genes, see Weintraub, H. *et al.* (1986) *Trends Genet.* 1:Review.

**[00170]** Another aspect the invention provides a host cell which includes a nucleic acid molecule described herein, e.g., a 93870 nucleic acid molecule within a recombinant expression vector or a 93870 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell, but also to the progeny or potential progeny of such a cell. Because certain modifications can occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are included within the scope of the term as used herein.

10085233.022802

**[00171]** A host cell can be any prokaryotic or eukaryotic cell. For example, a 93870 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary (CHO) cells) or COS cells. Other suitable host cells are known to those skilled in the art.

**[00172]** Vector DNA can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

**[00173]** A host cell of the invention can be used to produce (i.e., express) a 93870 protein. Accordingly, the invention further provides methods for producing a 93870 protein using the host cells of the invention. In one embodiment, the method includes culturing the host cell of the invention (into which a recombinant expression vector encoding a 93870 protein has been introduced) in a suitable medium such that a 93870 protein is produced. In another embodiment, the method further includes isolating a 93870 protein from the medium or the host cell.

**[00174]** In another aspect, the invention features, a cell or purified preparation of cells which include a 93870 transgene, or which otherwise mis-express 93870. The cell preparation can consist of human or non-human cells, e.g., rodent cells, e.g., mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a 93870 transgene, e.g., a heterologous form of a 93870, e.g., a gene derived from humans (in the

case of a non-human cell). The 93870 transgene can be mal-expressed, e.g., over-expressed or under-expressed. In other preferred embodiments, the cell or cells include a gene that mal-expresses an endogenous 93870, e.g., a gene the expression of which is disrupted, e.g., a knockout. Such cells can serve as a model for studying disorders that are related to mutated or mal-expressed 93870 alleles or for use in drug screening.

**[00175]** In another aspect, the invention includes, a human cell, e.g., a hematopoietic stem cell, transformed with nucleic acid that encodes a subject 93870 polypeptide.

**[00176]** Also provided are cells, preferably human cells, e.g., human hematopoietic or fibroblast cells, in which an endogenous 93870 is under the control of a regulatory sequence that does not normally control expression of the endogenous 93870 gene. The expression characteristics of an endogenous gene within a cell, e.g., a cell line or microorganism, can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the endogenous 93870 gene. For example, an endogenous 93870 gene that is "transcriptionally silent," e.g., not normally expressed, or expressed only at very low levels, can be activated by inserting a regulatory element that is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombination, can be used to insert the heterologous DNA as described (e.g., U.S. Patent number 5,272,071; PCT publication number WO 91/06667).

Transgenic Animals

**[00177]** The invention provides non-human transgenic animals. Such animals are useful for studying the function and/or activity of a 93870 protein and for identifying and/or evaluating modulators of 93870 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA or a rearrangement, e.g., a deletion of endogenous chromosomal DNA, which preferably is integrated into or occurs in the genome of the cells of a transgenic animal. A transgene can direct the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, other transgenes, e.g., a knockout, reduce expression. Thus, a transgenic animal can be one in which an endogenous 93870 gene has been altered, e.g., by homologous recombination

between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal (e.g., an embryonic cell of the animal, prior to development of the animal).

**[00178]** Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a transgene of the invention to direct expression of a 93870 protein to particular cells. A transgenic founder animal can be identified based upon the presence of a 93870 transgene in its genome and/or expression of 93870 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a 93870 protein can further be bred to other transgenic animals carrying other transgenes.

**[00179]** 93870 proteins or polypeptides can be expressed in transgenic animals or plants, e.g., a nucleic acid encoding the protein or polypeptide can be introduced into the genome of an animal. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, e.g., a milk- or egg-specific promoter, and recovered from the milk or eggs produced by the animal. Suitable animals are mice, pigs, cows, goats, and sheep.

**[00180]** The invention also includes a population of cells from a transgenic animal, as described herein.

10085233.022802

Uses

**[00181]** The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (e.g., therapeutic and prophylactic). The isolated nucleic acid molecules of the invention can be used, for example, to express a 93870 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect a 93870 mRNA (e.g., in a biological sample), to detect a genetic alteration in a 93870 gene and to modulate 93870 activity, as described further below. The 93870 proteins can be used to treat disorders characterized by insufficient or excessive production of a 93870 substrate or production of 93870 inhibitors. In addition, the 93870 proteins can be used to screen for naturally occurring 93870 substrates, to screen for drugs or compounds which modulate 93870 activity, as well as to treat disorders characterized by insufficient or excessive production of 93870 protein or production of 93870 protein forms

which have decreased, aberrant or unwanted activity compared to 93870 wild-type protein. Exemplary disorders include those in which G- protein coupled receptor activity is aberrant (e.g., congenital stationary night blindness, cardiomyopathy, and familial nephrogenic diabetes insipidus). Moreover, the anti-93870 antibodies of the invention can be used to detect and isolate 93870 proteins, regulate the bioavailability of 93870 proteins, and modulate 93870 activity.

**[00182]** A method of evaluating a compound for the ability to interact with, e.g., bind to, a subject 93870 polypeptide is provided. The method includes: contacting the compound with the subject 93870 polypeptide; and evaluating the ability of the compound to interact with, e.g., to bind or form a complex with, the subject 93870 polypeptide. This method can be performed *in vitro*, e.g., in a cell free system, or *in vivo*, e.g., in a two-hybrid interaction trap assay. This method can be used to identify naturally-occurring molecules that interact with a subject 93870 polypeptide. It can also be used to find natural or synthetic inhibitors of a subject 93870 polypeptide. Screening methods are discussed in more detail below.

<u>Screening Assays</u>

**[00183]** The invention provides screening methods (also referred to herein as "assays") for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind with 93870 proteins, have a stimulatory or inhibitory effect on, for example, 93870 expression or 93870 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a 93870 substrate. Compounds thus identified can be used to modulate the activity of target gene products (e.g., 93870 genes) in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions.

**[00184]** In one embodiment, the invention provides assays for screening candidate or test compounds that are substrates of a 93870 protein or polypeptide or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds that bind to or modulate the activity of a 93870 protein or polypeptide or a biologically active portion thereof.

**[00185]** The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of



10085233.022802

peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive; e.g., Zuckermann *et al.* (1994) *J. Med. Chem.* 37:2678-2685); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, 1997, Anticancer Drug Des. 12:145).

**[00186]** Examples of methods for the synthesis of molecular libraries have been described (e.g., DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994) *J. Med. Chem.* 37:2678; Cho *et al,* (1993) *Science* 261:1303; Carrell *et al.* (1994) Angew. *Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) Angew. *Chem. Int. Ed. Engl.* 33:2061; and Gallop *et al.*, (1994) *J. Med. Chem.* 37:1233).

**[00187]** Libraries of compounds can be presented in solution (e.g., Houghten, (1992) *Biotechniques* 13:412-421), or on beads Lam (1991) *Nature* 354:82-84), chips (Fodor, 1993, *Nature* 364:555-556), bacteria (U.S. Patent number 5,223,409), spores (U.S. Patent number 5,223,409), plasmids (Cull *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869), or on phage (Scott *et al.* (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:6378-6382; Felici (1991) *J. Mol. Biol.* 222:301-310; U.S. Patent number 5,223,409).

**[00188]** In one embodiment, an assay is a cell-based assay in which a cell which expresses a 93870 protein or biologically active portion thereof is contacted with a test compound, and the ability of the test compound to modulate 93870 activity is determined. Determining the ability of the test compound to modulate 93870 activity can be accomplished by monitoring, for example, changes in enzymatic activity. The cell, for example, can be of mammalian origin.

**[00189]** The ability of the test compound to modulate 93870 binding to a compound, e.g., a 93870 substrate, or to bind to 93870 can also be evaluated. This can be accomplished, for example, by coupling the compound, e.g., the substrate, with a radioisotope or enzymatic label such that binding of the compound, e.g., the substrate, to 93870 can be determined by detecting the labeled compound, e.g., substrate, in a complex. Alternatively, 93870 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to

modulate 93870 binding to a 93870 substrate in a complex. For example, compounds (e.g., 93870 substrates) can be labeled with $^{125}I$, $^{35}S$, $^{14}C$, or $^{3}H$, either directly or indirectly, and the radioisotope detected by direct counting of radio-emission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

**[00190]** The ability of a compound (e.g., a 93870 substrate) to interact with 93870 with or without the labeling of any of the interactants can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with 93870 without the labeling of either the compound or the 93870 (McConnell *et al.* (1992,) *Science* 257:1906-1912). As used herein, a "microphysiometer" (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and 93870.

**[00191]** In yet another embodiment, a cell-free assay is provided in which a 93870 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the 93870 protein or biologically active portion thereof is evaluated. Preferred biologically active portions of the 93870 proteins to be used in assays of the present invention include fragments that participate in interactions with non-93870 molecules, e.g., fragments with high surface probability scores.

**[00192]** Soluble and/or membrane-bound forms of isolated proteins (e.g., 93870 proteins or biologically active portions thereof) can be used in the cell-free assays of the invention. When membrane-bound forms of the protein are used, it can be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)n, 3-{(3-cholamidopropyl) dimethylamminio}-1-propane sulfonate (CHAPS), 3-{(3-cholamidopropyl) dimethylamminio}-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

**[00193]** Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two



10085233.022802

components to interact and bind, thus forming a complex that can be removed and/or detected.

**[00194]** The interaction between two molecules can also be detected, e.g., using fluorescence energy transfer (FET; e.g., U.S. Patent number 5,631,169; U.S. Patent number 4,868,103). A fluorophore label is selected such that a first donor molecule's emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule can simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label can be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

**[00195]** In another embodiment, determining the ability of the 93870 protein to bind to a target molecule can be accomplished using real-time biomolecular interaction analysis (BIA; e.g., Sjolander *et al.* (1991) *Anal. Chem.* 63:2338-2345; Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705). "Surface plasmon resonance" (SPR) or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of SPR), resulting in a detectable signal that can be used as an indication of real-time reactions between biological molecules.

**[00196]** In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

**[00197]** It can be desirable to immobilize either 93870, an anti-93870 antibody or its target molecule to facilitate separation of complexed from non-complexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a 93870 protein, or interaction of a 93870 protein with a target molecule in the

presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/93870 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione Sepharose™ beads (Sigma Chemical, St. Louis, MO) or glutathione-derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or 93870 protein, and the mixture incubated under conditions conducive for complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of 93870 binding or activity determined using standard techniques.

**[00198]** Other techniques for immobilizing either a 93870 protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated 93870 protein or target molecules can be prepared from biotin- N-hydroxy-succinimide using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

**[00199]** In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, non-reacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

**[00200]** In one embodiment, this assay is performed utilizing antibodies reactive with 93870 protein or target molecules but which do not interfere with binding of the 93870 protein to its target molecule. Such antibodies can be derivatized to the wells of the plate,

and unbound target or 93870 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the 93870 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the 93870 protein or target molecule.

**[00201]** Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from non-reacted components, by any of a number of standard techniques, including, but not limited to: differential centrifugation (e.g., Rivas *et al.* (1993) *Trends Biochem. Sci.* 18:284-287); chromatography (e.g., gel filtration chromatography or ion-exchange chromatography); electrophoresis (e.g., Ausubel *et al.* eds. (1999) *Current Protocols in Molecular Biology*, J. Wiley, New York); and immunoprecipitation (e.g., Ausubel, *supra*). Such resins and chromatographic techniques are known to one skilled in the art (e.g., Heegaard, 1998, J. Mol. Recognit. 11:141-148; Hage *et al.* (1997) *J. Chromatogr. B Biomed. Sci. Appl.* 699:499-525). Further, fluorescence energy transfer can also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

**[00202]** In a preferred embodiment, the assay includes contacting the 93870 protein or biologically active portion thereof with a known compound which binds 93870 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a 93870 protein, wherein determining the ability of the test compound to interact with a 93870 protein includes determining the ability of the test compound to preferentially bind to 93870 or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

**[00203]** The target gene products of the invention can, *in vivo*, interact with one or more cellular or extracellular macromolecules, such as proteins. For the purposes of this discussion, such cellular and extracellular macromolecules are referred to herein as "binding partners." Compounds that disrupt such interactions can be useful in regulating the activity of the target gene product. Such compounds can include, but are not limited to molecules such as antibodies, peptides, and small molecules. The preferred target genes/products for use in this embodiment are the 93870 genes herein identified. In an alternative embodiment, the invention provides methods for determining the ability of the test compound to modulate the activity of a 93870 protein through modulation of the activity of a downstream effector of a 93870 target molecule. For example, the activity of the effector molecule on an



10085233 .022802

appropriate target can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

**[00204]** To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), a reaction mixture containing the target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

**[00205]** These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding partner onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the target gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described below.

**[00206]** In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (e.g., a microtiter

10085233.022802

plate), while the non-anchored species is labeled, either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the species to be anchored can be used to anchor the species to the solid surface.

**[00207]** In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, non-reacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

10085233.022802

**[00208]** Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from non-reacted components, and complexes detected; e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

**[00209]** In an alternate embodiment of the invention, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (e.g., U.S. Patent number 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

**[00210]** In yet another aspect, the 93870 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (e.g., U.S. Patent number 5,283,317; Zervos *et al.*, (1993)

*Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; PCT publication number WO 94/10300), to identify other proteins, which bind to or interact with 93870 ("93870-binding proteins" or "93870-bp") and are involved in 93870 activity. Such 93870-bps can be activators or inhibitors of signals by the 93870 proteins or 93870 targets as, for example, downstream elements of a 93870-mediated signaling pathway.

**[00211]** The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a 93870 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively, the 93870 protein can be fused to the activator domain). If the "bait" and the "prey" proteins are able to interact *in vivo* forming a 93870-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein that interacts with the 93870 protein.

**[00212]** In another embodiment, modulators of 93870 expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of 93870 mRNA or protein evaluated relative to the level of expression of 93870 mRNA or protein in the absence of the candidate compound. When expression of 93870 mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of 93870 mRNA or protein expression. Alternatively, when expression of 93870 mRNA or protein is less (i.e., statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 93870 mRNA or protein expression. The level of 93870 mRNA or protein expression can be determined by methods described herein for detecting 93870 mRNA or protein.

**[00213]** In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-

10085233.022802

based or a cell free assay, and the ability of the agent to modulate the activity of a 93870 protein can be confirmed *in vivo*, e.g., in an animal such as an animal model for a disease.

**[00214]** This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein (e.g., a 93870 modulating agent, an antisense 93870 nucleic acid molecule, a 93870-specific antibody, or a 93870-binding partner) in an appropriate animal model to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-described screening assays can be used for treatments as described herein.

Detection Assays

**[00215]** Portions or fragments of the nucleic acid sequences identified herein can be used as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome, e.g., to locate gene regions associated with genetic disease or to associate 93870 with a disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

Chromosome Mapping

**[00216]** The 93870 nucleotide sequences or portions thereof can be used to map the location of the 93870 genes on a chromosome. This process is called chromosome mapping. Chromosome mapping is useful in correlating the 93870 sequences with genes associated with disease.

**[00217]** Briefly, 93870 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 base pairs in length) from the 93870 nucleotide sequence (e.g., SEQ ID NO:1 or SEQ ID NO:3). These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the 93870 sequences will yield an amplified fragment.

**[00218]** A panel of somatic cell hybrids in which each cell line contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, can allow easy mapping of individual genes to specific human chromosomes (D'Eustachio *et al.* (1983) *Science* 220:919-924).



10085233.022802

**[00219]** Other mapping strategies e.g., *in situ* hybridization as described (Fan *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-6227), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries can be used to map 93870 to a chromosomal location.

**[00220]** Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of FISH, (see Verma *et al.* (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York).

**[00221]** Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to non-coding regions of the genes are typically preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

**[00222]** Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data (such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), as described (e.g., Egeland *et al.*, 1987, Nature, 325:783-787).

**[00223]** Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the 93870 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete

10085233.022802

sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

**[00224]** 93870 sequences can be used to identify individuals from biological samples using, e.g., restriction fragment length polymorphism (RFLP). In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, the fragments separated, e.g., in a Southern blot, and probed to yield bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent number 5,272,057).

**[00225]** Furthermore, the sequences of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the 93870 nucleotide sequence described herein can be used to prepare PCR primers homologous to the 5'- and 3'-ends of the sequence. These primers can then be used to amplify an individual's DNA and subsequently sequence it. Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences.

10085233.022802

**[00226]** Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the non-coding regions. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the non-coding regions, fewer sequences are necessary to differentiate individuals. The non-coding sequences of SEQ ID NO:1 can provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a non-coding amplified sequence of 100 bases. If predicted coding sequences are used, such as those in SEQ ID NO:3, a more appropriate number of primers for positive individual identification would be 500-2,000.

**[00227]** If a panel of reagents from 93870 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

Use of Partial 93870 Sequences in Forensic Biology

**[00228]** DNA-based identification techniques can also be used in forensic biology. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

**[00229]** The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e., another DNA sequence that is unique to a particular individual). As mentioned above, actual nucleotide sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to non-coding regions of SEQ ID NO:1 (e.g., fragments having a length of at least 20 nucleotide residues, preferably at least 30 nucleotide residues) are particularly appropriate for this use.

**[00230]** The 93870 nucleotide sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or label-able probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., a tissue containing hematopoietic cells. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such 93870 probes can be used to identify tissue by species and/or by organ type.

**[00231]** In a similar fashion, these reagents, e.g., 93870 primers or probes can be used to screen tissue culture for contamination (i.e., to screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine

**[00232]** The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual.

**[00233]** Generally, the invention provides a method of determining if a subject is at risk for a disorder related to a lesion in, or the mis-expression of, a gene that encodes a 93870 polypeptide.



10085253 .022802

**[00234]** Such disorders include, e.g., a disorder associated with the mis-expression of a 93870 polypeptide, e.g., an immune disorder or a neoplastic disorder.

**[00235]** The method includes one or more of the following:

**[00236]** detecting, in a tissue of the subject, the presence or absence of a mutation which affects the expression of the 93870 gene, or detecting the presence or absence of a mutation in a region which controls the expression of the gene, e.g., a mutation in the 5'-control region;

**[00237]** detecting, in a tissue of the subject, the presence or absence of a mutation which alters the structure of the 93870 gene;

**[00238]** detecting, in a tissue of the subject, the mis-expression of the 93870 gene at the mRNA level, e.g., detecting a non-wild-type level of a mRNA; and

**[00239]** detecting, in a tissue of the subject, the mis-expression of the gene at the protein level, e.g., detecting a non-wild-type level of a 93870 polypeptide.

**[00240]** In preferred embodiments the method includes: ascertaining the existence of at least one of: a deletion of one or more nucleotides from the 93870 gene; an insertion of one or more nucleotides into the gene, a point mutation, e.g., a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, e.g., a translocation, inversion, or deletion.

**[00241]** For example, detecting the genetic lesion can include: (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence from SEQ ID NO:1, or naturally occurring mutants thereof, or 5'- or 3'-flanking sequences naturally associated with the 93870 gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and detecting the presence or absence of the genetic lesion by hybridization of the probe/primer to the nucleic acid, e.g., by *in situ* hybridization.

**[00242]** In preferred embodiments, detecting the mis-expression includes ascertaining the existence of at least one of: an alteration in the level of a messenger RNA transcript of the 93870 gene; the presence of a non-wild-type splicing pattern of a messenger RNA transcript of the gene; or a non-wild-type level of 93870 RNA or protein.

**[00243]** Methods of the invention can be used for prenatal screening or to determine if a subject's offspring will be at risk for a disorder.

**[00244]** In preferred embodiments the method includes determining the structure of a 93870 gene, an abnormal structure being indicative of risk for the disorder.



10085233.022802

**[00245]** In preferred embodiments the method includes contacting a sample form the subject with an antibody to the 93870 protein or a nucleic acid, which hybridizes specifically with the gene. These and other embodiments are discussed below.

Diagnostic and Prognostic Assays

**[00246]** The presence, level, or absence of 93870 protein or nucleic acid in a biological sample can be evaluated by obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 93870 protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes 93870 protein such that the presence of 93870 protein or nucleic acid is detected in the biological sample. The term "biological sample" includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. A preferred biological sample is serum. The level of expression of the 93870 gene can be measured in a number of ways, including, but not limited to: measuring the mRNA encoded by the 93870 genes; measuring the amount of protein encoded by the 93870 genes; or measuring the activity of the protein encoded by the 93870 genes.

**[00247]** The level of mRNA corresponding to the 93870 gene in a cell can be determined both by *in situ* and by *in vitro* formats.

**[00248]** The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a full-length 93870 nucleic acid, such as the nucleic acid of SEQ ID NO:1, the deposited nucleotide sequence, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to 93870 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays are described herein.

**[00249]** In one format, mRNA (or cDNA) is immobilized on a surface and contacted with the probes, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probes are immobilized on a surface and the mRNA (or cDNA) is contacted with the probes, for example, in a two-dimensional gene chip array. A skilled artisan can adapt known



10085233.022802

mRNA detection methods for use in detecting the level of mRNA encoded by the 93870 genes.

**[00250]** The level of mRNA in a sample that is encoded by 93870 can be evaluated with nucleic acid amplification, e.g., by RT-PCR (U.S. Patent number 4,683,202), ligase chain reaction (Barany, 1991, Proc. Natl. Acad. Sci. USA 88:189-193), self-sustained sequence replication (Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.* (1988) *Bio/Technology* 6:1197), rolling circle replication (U.S. Patent number 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques known in the art. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5'- or 3'-regions of a 93870 gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence between the primers.

**[00251]** For *in situ* methods, a cell or tissue sample can be prepared/processed and immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the 93870 gene being analyzed.

**[00252]** In another embodiment, the methods include further contacting a control sample with a compound or agent capable of detecting 93870 mRNA, or genomic DNA, and comparing the presence of 93870 mRNA or genomic DNA in the control sample with the presence of 93870 mRNA or genomic DNA in the test sample.

**[00253]** A variety of methods can be used to determine the level of protein encoded by 93870. In general, these methods include contacting an agent that selectively binds to the protein, such as an antibody with a sample, to evaluate the level of protein in the sample. In a preferred embodiment, the antibody bears a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be used. The term "labeled," with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with a detectable substance. Examples of detectable substances are provided herein.



10085233.022802

**[00254]** The detection methods can be used to detect 93870 protein in a biological sample *in vitro* as well as *in vivo*. *In vitro* techniques for detection of 93870 protein include enzyme linked immunosorbent assays (ELISAs), immunoprecipitations, immunofluorescence, enzyme immunoassay (EIA), radioimmunoassay (RIA), and Western blot analysis. *In vivo* techniques for detection of 93870 protein include introducing into a subject a labeled anti-93870 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

**[00255]** In another embodiment, the methods further include contacting the control sample with a compound or agent capable of detecting 93870 protein, and comparing the presence of 93870 protein in the control sample with the presence of 93870 protein in the test sample.

**[00256]** The invention also includes kits for detecting the presence of 93870 in a biological sample. For example, the kit can include a compound or agent capable of detecting 93870 protein or mRNA in a biological sample, and a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 93870 protein or nucleic acid.

**[00257]** For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide corresponding to a marker of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

**[00258]** For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, e.g., a detectably-labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also includes a buffering agent, a preservative, or a protein-stabilizing agent. The kit can also includes components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples that can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

**[00259]** The diagnostic methods described herein can identify subjects having, or at risk of developing, a disease or disorder associated with misexpressed, aberrant or unwanted

93870 expression or activity. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as pain or deregulated cell proliferation.

**[00260]** In one embodiment, a disease or disorder associated with aberrant or unwanted 93870 expression or activity is identified. A test sample is obtained from a subject and 93870 protein or nucleic acid (e.g., mRNA or genomic DNA) is evaluated, wherein the level, e.g., the presence or absence, of 93870 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted 93870 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest, including a biological fluid (e.g., serum), cell sample, or tissue.

**[00261]** The prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted 93870 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent that modulates 93870 expression or activity.

**[00262]** The methods of the invention can also be used to detect genetic alterations in a 93870 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in 93870 protein activity or nucleic acid expression, such as a disorder associated with chemokine interactions, inflammation, neurotransmission, light reception, or hormone reception. In preferred embodiments, the methods include detecting, in a sample from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a 93870 protein, or the mis-expression of the 93870 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a 93870 gene; 2) an addition of one or more nucleotides to a 93870 gene; 3) a substitution of one or more nucleotides of a 93870 gene, 4) a chromosomal rearrangement of a 93870 gene; 5) an alteration in the level of a messenger RNA transcript of a 93870 gene, 6) aberrant modification of a 93870 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a 93870 gene, 8) a non-wild-type level of a 93870 protein, 9) allelic loss of a 93870 gene, and 10) inappropriate post-translational modification of a 93870 protein.

**[00263]** An alteration can be detected without a probe/primer in a polymerase chain reaction, such as anchor PCR or RACE-PCR, or, alternatively, in a ligation chain reaction (LCR), the latter of which can be particularly useful for detecting point mutations in the 93870 gene. This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (e.g., genomic, mRNA or both) from the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a 93870 gene under conditions such that hybridization and amplification of the 93870 gene occurs (if present), and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR can be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

**[00264]** Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.* (1988) *Bio/Technology* 6:1197), or other nucleic acid amplification methods, followed by the detection of the amplified molecules using techniques known to those of skill in the art.

**[00265]** In another embodiment, mutations in a 93870 gene from a sample cell can be identified by detecting alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined, e.g., by gel electrophoresis, and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (e.g., U.S. Patent number 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

**[00266]** In other embodiments, genetic mutations in 93870 can be identified by hybridizing a sample to control nucleic acids, e.g., DNA or RNA, by, e.g., two-dimensional arrays, or, e.g., chip based arrays. Such arrays include a plurality of addresses, each of which is positionally distinguishable from the other. A different probe is located at each address of the plurality. The arrays can have a high density of addresses, e.g., can contain hundreds or thousands of oligonucleotides probes (Cronin *et al.* (1996) *Hum. Mutat.* 7:244-255; Kozal *et al.* (1996) *Nature Med.* 2:753-759). For example, genetic mutations in 93870 can be identified in two-dimensional arrays containing light-generated DNA probes as described

10085233.022802

(Cronin *et al.supra*). Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

**[00267]** In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 93870 gene and detect mutations by comparing the sequence of the sample 93870 with the corresponding wild-type (control) sequence. Automated sequencing procedures can be utilized when performing the diagnostic assays (1995, Biotechniques 19:448), including sequencing by mass spectrometry.

**[00268]** Other methods for detecting mutations in the 93870 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242; Cotton *et al.*, (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba *et al,* (1992) *Meth. Enzymol.* 217:286-295).

**[00269]** In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in 93870 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994*) Carcinogenesis* 15:1657-1662; U.S. Patent number 5,459,039).

**[00270]** In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 93870 genes. For example, single strand conformation polymorphism (SSCP) can be used to detect differences in electrophoretic mobility between mutant and wild-type nucleic acids (Orita *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:2766; Cotton (1993) *Mutat. Res.* 285:125-144; Hayashi (1992) *Genet. Anal.* Tech. Appl. 9:73-79). Single-stranded DNA fragments of sample and control 93870 nucleic acids will be denatured and allowed to re-nature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments can be labeled or detected with

labeled probes. The sensitivity of the assay can be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

**[00271]** In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 base pairs of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) Biophys Chem 265:12753).

**[00272]** Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension (Saiki *et al* (1986) *Nature* 324:163; Saiki *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6230).

**[00273]** Alternatively, allele specific amplification technology that depends on selective PCR amplification can be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification can carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; Gibbs *et al.* (1989) *Nucl. Acids Res.* 17:2437-2448) or at the extreme 3'-end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner, 1993, Tibtech 11:238). In addition, it can be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments, amplification can also be performed using Taq ligase for amplification (Barany, 1991, Proc. Natl. Acad. Sci. USA 88:189). In such cases, ligation will occur only if there is a perfect match at the 3'-end of the 5'- sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

**[00274]** The methods described herein can be performed, for example, using pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which can be conveniently used, e.g., in clinical settings to diagnose

10085233.022802

patients exhibiting symptoms or family history of a disease or illness involving a 93870 gene.

Use of 93870 Molecules as Surrogate Markers

**[00275]** The 93870 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the 93870 molecules of the invention can be detected, and can be correlated with one or more biological states *in vivo*. For example, the 93870 molecules of the invention can serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (e.g., with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers can serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (e.g., early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (e.g., an assessment of cardiovascular disease can be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection can be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers have been described (e.g., Koomen *et al.* (2000) *J. Mass. Spectrom.* 35:258-264; James (1994) *AIDS Treat. News Arch.* 209).

**[00276]** The 93870 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker can be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to

10085233.022802

the level of the drug. In this fashion, the distribution or uptake of the drug can be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker can be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug can be sufficient to activate multiple rounds of marker (e.g., a 93870 marker) transcription or expression, the amplified marker can be in a quantity which is more readily detectable than the drug itself. Also, the marker can be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-93870 antibodies can be employed in an immune-based detection system for a 93870 protein marker, or 93870-specific radiolabeled probes can be used to detect a 93870 mRNA marker. Furthermore, the use of a pharmacodynamic marker can offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers have been described (e.g., U.S. Patent number 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90: 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; Nicolau (1999) *Am, J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20).

10085233.022802

**[00277]** The 93870 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (e.g., McLeod *et al.* (1999) *Eur. J. Cancer* 35:1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, can be selected. For example, based on the presence or quantity of RNA, or protein (e.g., 93870 protein or RNA) for specific tumor markers in a subject, a drug or course of treatment can be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in 93870 DNA can correlate 93870 drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

Pharmaceutical Compositions

**[00278]** The nucleic acid and polypeptides, fragments thereof, as well as anti-93870 antibodies and small molecule modulators of 93870 (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions. Such compositions typically include the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

**[00279]** A pharmaceutical composition is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

**[00280]** Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example,

glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including an agent in the composition that delays absorption, for example, aluminum monostearate and gelatin.

**[00281]** Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying, which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

10035233.022802

**[00282]** Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, e.g., gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder, such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient, such as starch or lactose; a disintegrating agent, such as alginic acid, Primogel™, or corn starch; a lubricant, such as magnesium stearate or Sterotes™; a glidant, such as colloidal silicon dioxide; a sweetening agent, such as sucrose or saccharin; or a flavoring agent, such as peppermint, methyl salicylate, or orange flavoring.

**[00283]** For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser that contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

**[00284]** Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

**[00285]** The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

10085233.022802

**[00286]** In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells using monoclonal antibodies directed towards viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to described methods (e.g., U.S. Patent number 4,522,811).

**[00287]** It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

**[00288]** Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the $LD_{50}$ (the dose lethal to 50% of the population) and the $ED_{50}$ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio $LD_{50}/ED_{50}$.

Compounds that exhibit high therapeutic indices are preferred. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

**[00289]** The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the $ED_{50}$ with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the $IC_{50}$ (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

10085233.022802

**[00290]** As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 milligrams per kilogram body weight, preferably about 0.01 to 25 milligrams per kilogram body weight, more preferably about 0.1 to 20 milligrams per kilogram body weight, and even more preferably about 1 to 10 milligrams per kilogram, 2 to 9 milligrams per kilogram, 3 to 8 milligrams per kilogram, 4 to 7 milligrams per kilogram, or 5 to 6 milligrams per kilogram body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors can influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

**[00291]** For antibodies, the preferred dosage is 0.1 milligrams per kilogram of body weight (generally 10 to 20 milligrams per kilogram). If the antibody is to act in the brain, a dosage of 50 to 100 milligrams per kilogram is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body

than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for the lipidation of antibodies is described by Cruikshank *et al.* (1997) *J. AIDS Hum. Retrovir.* 14:193.

**[00292]** The present invention encompasses agents that modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including hetero-organic and organo-metallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

10085233.022802

**[00293]** Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

**[00294]** An antibody (or fragment thereof) can be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine,

vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

**[00295]** The conjugates of the invention can be used for modifying a given biological response, and the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety can be a protein or polypeptide possessing a desired biological activity. Such proteins can include, for example, a toxin such as abrin, ricin A, gelonin, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukins-1, -2, and -6, granulocyte macrophage colony stimulating factor, granulocyte colony stimulating factor, or other growth factors.

**[00296]** Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent number 4,676,980.

**[00297]** The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent number 5,328,470) or by stereotactic injection (e.g., Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

**[00298]** The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

10085233.022802

Methods of Treatment

**[00299]** The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted 93870 expression or activity. With regards to both prophylactic and therapeutic methods of treatment, such treatments can be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics," as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype," or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 93870 molecules of the present invention or 93870 modulators according to that individual's drug response genotype.

**[00300]** Treatment is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease.

**[00301]** A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

**[00302]** Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

**[00303]** In one aspect, the invention provides a method for preventing a disease or condition in a subject associated with an aberrant or unwanted 93870 expression or activity, by administering to the subject a 93870 or an agent which modulates 93870 expression, or at least one 93870 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted 93870 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the 93870 aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its



10085233.022802

progression. Depending on the type of 93870 aberrance, for example, a 93870 agonist or 93870 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

**[00304]** It is possible that some 93870 disorders can be caused, at least in part, by an abnormal level of gene product, or by the presence of a gene product exhibiting abnormal activity. As such, the reduction in the level and/or activity of such gene products would bring about the amelioration of disorder symptoms.

**[00305]** As discussed, successful treatment of 93870 disorders can be brought about by techniques that serve to inhibit the expression or activity of target gene products. For example, compounds, e.g., an agent identified using an assays described above, that proves to exhibit negative modulatory activity, can be used in accordance with the invention to prevent and/or ameliorate symptoms of 93870 disorders. Such molecules can include, but are not limited to peptides, phosphopeptides, small organic or inorganic molecules, or antibodies (including, for example, polyclonal, monoclonal, human, humanized, anti-idiotypic, chimeric or single chain antibodies, and Fab, $F(ab')_2$ and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof).

**[00306]** Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used in accordance with the invention to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

**[00307]** It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular protein, it can be preferable to co-administer normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

**[00308]** Another method by which nucleic acid molecules can be utilized in treating or preventing a disease characterized by 93870 expression is through the use of aptamer

molecules specific for 93870 protein. Aptamers are nucleic acid molecules having a tertiary structure that permits them to specifically bind to protein ligands (e.g., Osborne *et al.*, (1997) *Curr. Opin. Chem. Biol.* 1:5-9; Patel (1997) *Curr. Opin. Chem.* Biol. 1:32-46). Since nucleic acid molecules can in many cases be more conveniently introduced into target cells than therapeutic protein molecules can be, aptamers offer a method by which 93870 protein activity can be specifically decreased without the introduction of drugs or other molecules which can have pluripotent effects.

**[00309]** Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies may, therefore, by administered in instances whereby negative modulatory techniques are appropriate for the treatment of 93870 disorders.

**[00310]** In circumstances wherein injection of an animal or a human subject with a 93870 protein or epitope for stimulating antibody production is harmful to the subject, it is possible to generate an immune response against 93870 through the use of anti-idiotypic antibodies (e.g., Herlyn (1999) *Ann. Med.* 31:66-78; Bhattacharya-Chatterjee *et al.* (1998) *Cancer Treat. Res.* 94:51-68. If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the 93870 protein. Vaccines directed to a disease characterized by 93870 expression can also be generated in this fashion.

**[00311]** In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies can be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (e.g., Marasco *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:7889-7893).

**[00312]** The identified compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent, treat or ameliorate GPCR disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders.



10085233.022802

**[00313]** Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the $LD_{50}$ (the dose lethal to 50% of the population) and the $ED_{50}$ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio $LD_{50}/ED_{50}$. Compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

**[00314]** The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the $ED_{50}$ with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the $IC_{50}$ (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

**[00315]** Another example of determination of effective dose for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays can utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. The compound which is able to modulate 93870 activity is used as a template, or "imprinting molecule," to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix that contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. Detailed reviews of this technique appear in the art (Ansell *et al.* (1996) *Curr. Opin. Biotechnol.* 7:89-94; Shea (1994) Trends Polymer Sci. 2:166-173). Such "imprinted" affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix (e.g., a matrix described in Vlatakis *et al.* (1993) *Nature* 361:645-647. Through the use of isotope-

10085233.022802

labeling, the "free" concentration of compound which modulates the expression or activity of 93870 can be readily monitored and used in calculations of $IC_{50}$.

**[00316]** Such "imprinted" affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using appropriate fiber optic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual $IC_{50}$. A rudimentary example of such a "biosensor" is discussed in Kriz *et al.* (1995) *Anal. Chem.* 67:2142-2144.

**[00317]** Another aspect of the invention pertains to methods of modulating 93870 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a 93870 or agent that modulates one or more of the activities of 93870 protein activity associated with the cell. An agent that modulates 93870 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a 93870 protein (e.g., a 93870 substrate or receptor), a 93870 antibody, a 93870 agonist or antagonist, a peptidomimetic of a 93870 agonist or antagonist, or other small molecule.

10085233.022802

**[00318]** In one embodiment, the agent stimulates one or 93870 activities. Examples of such stimulatory agents include active 93870 protein and a nucleic acid molecule encoding 93870. In another embodiment, the agent inhibits one or more 93870 activities. Examples of such inhibitory agents include antisense 93870 nucleic acid molecules, anti-93870 antibodies, and 93870 inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a 93870 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) 93870 expression or activity. In another embodiment, the method involves administering a 93870 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted 93870 expression or activity.

**[00319]** Stimulation of 93870 activity is desirable in situations in which 93870 is abnormally down-regulated and/or in which increased 93870 activity is likely to have a beneficial effect. For example, stimulation of 93870 activity is desirable in situations in

which a 93870 is down-regulated and/or in which increased 93870 activity is likely to have a beneficial effect. Likewise, inhibition of 93870 activity is desirable in situations in which 93870 is abnormally up-regulated and/or in which decreased 93870 activity is likely to have a beneficial effect.

**[00320]** The 93870 molecules can act as novel diagnostic targets and therapeutic agents for controlling one or more of immune and inflammatory disorders, platelet disorders, skeletal or bone metabolism disorders, and bone marrow mononuclear cell disorders described above, as well as cellular proliferative and/or differentiative disorders, hormonal disorders, neurological disorders, cardiovascular disorders, blood vessel disorders, viral diseases, liver disorders, and pain and metabolic disorders described below.

**[00321]** Examples of cellular proliferative and/or differentiative disorders include cancer, e.g., carcinoma, sarcoma, metastatic disorders or hematopoietic neoplastic disorders, e.g., leukemias. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited to those of prostate, colon, lung, breast and liver origin.

**[00322]** As used herein, the term "cancer" (also used interchangeably with the terms, "hyperproliferative" and "neoplastic") refers to cells having the capacity for autonomous growth, i.e., an abnormal state or condition characterized by rapidly proliferating cell growth. Cancerous disease states may be categorized as pathologic, i.e., characterizing or constituting a disease state, e.g., malignant tumor growth, or may be categorized as non-pathologic, i.e., a deviation from normal but not associated with a disease state, e.g., cell proliferation associated with wound repair. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. The term "cancer" includes malignancies of the various organ systems, such as those affecting lung, breast, thyroid, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus. The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term "carcinoma"

also includes carcinosarcomas, e.g., which include malignant tumors composed of carcinomatous and sarcomatous tissues. An “adenocarcinoma” refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures. The term “sarcoma” is art recognized and refers to malignant tumors of mesenchymal derivation.

**[00323]** The 93870 molecules of the invention can be used to monitor, treat and/or diagnose a variety of proliferative disorders. Such disorders include hematopoietic neoplastic disorders. As used herein, the term “hematopoietic neoplastic disorders” includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, e.g., arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Typically, the diseases arise from poorly differentiated acute leukemias, e.g., erythroblastic leukemia and acute megakaryoblastic leukemia. Additional exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus, L., (1991) *Crit. Rev. in Oncol./Hemotol.* 11:267-97); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGF), Hodgkin's disease and Reed-Sternberg disease.

**[00324]** GPCR-associated disorders can include hormonal disorders, such as conditions or diseases in which the production and/or regulation of hormones in an organism is aberrant. Examples of such disorders and diseases include type I and type II diabetes mellitus, pituitary disorders (*e.g.*, growth disorders), thyroid disorders (*e.g.*, hypothyroidism or hyperthyroidism), and reproductive or fertility disorders (*e.g.*, disorders which affect the organs of the reproductive system, *e.g.*, the prostate gland, the uterus, or the vagina; disorders which involve an imbalance in the levels of a reproductive hormone in a subject; disorders affecting the ability of a subject to reproduce; and disorders affecting secondary sex characteristic development, *e.g.*, adrenal hyperplasia).

**[00325]** Additional GPCR-associated disorders are neurological disorders. Such neurological disorders include, for example, disorders involving neurons, and disorders



10085233.022802

involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; cerebrovascular diseases, such as those related to hypoxia, ischemia, and infarction, including hypotension, hypoperfusion, and low-flow states--global cerebral ischemia and focal cerebral ischemia--infarction from obstruction of local blood supply, intracranial hemorrhage, including intracerebral (intraparenchymal) hemorrhage, subarachnoid hemorrhage and ruptured berry aneurysms, and vascular malformations, hypertensive cerebrovascular disease, including lacunar infarcts, slit hemorrhages, and hypertensive encephalopathy; infections, such as acute meningitis, including acute pyogenic (bacterial) meningitis and acute aseptic (viral) meningitis, acute focal suppurative infections, including brain abscess, subdural empyema, and extradural abscess, chronic bacterial meningoencephalitis, including tuberculosis and mycobacterioses, neurosyphilis, and neuroborreliosis (Lyme disease), viral meningoencephalitis, including arthropod-borne (Arbo) viral encephalitis, *Herpes simplex* virus Type 1, *Herpes simplex* virus Type 2, *Varicella-zoster* virus (*Herpes zoster*), cytomegalovirus, poliomyelitis, rabies, and human immunodeficiency virus 1, including HIV-1 meningoencephalitis (subacute encephalitis), vacuolar myelopathy, AIDS-associated myopathy, peripheral neuropathy, and AIDS in children, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, fungal meningoencephalitis, other infectious diseases of the nervous system; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including multiple sclerosis, multiple sclerosis variants, acute disseminated encephalomyelitis and acute necrotizing hemorrhagic encephalomyelitis, and other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, including Alzheimer's disease and Pick's disease, degenerative diseases of basal ganglia and brain stem, including Parkinsonism, idiopathic Parkinson's disease (paralysis agitans), progressive supranuclear palsy, corticobasal degenration, multiple system atrophy, including striatonigral degenration, Shy-Drager syndrome, and olivopontocerebellar atrophy, and Huntington's disease; spinocerebellar degenerations, including spinocerebellar ataxias, including Friedreich ataxia, and ataxia-telanglectasia, degenerative diseases affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy; inborn errors of metabolism, such as

10085233.022802

leukodystrophies, including Krabbe disease, metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, and Canavan disease, mitochondrial encephalomyopathies, including Leigh disease and other mitochondrial encephalomyopathies; toxic and acquired metabolic diseases, including vitamin deficiencies such as thiamine (vitamin $B_1$) deficiency and vitamin $B_{12}$ deficiency, neurologic sequelae of metabolic disturbances, including hypoglycemia, hyperglycemia, and hepatic encephatopathy, toxic disorders, including carbon monoxide, methanol, ethanol, and radiation, including combined methotrexate and radiation-induced injury; tumors, such as gliomas, including astrocytoma, including fibrillary (diffuse) astrocytoma and glioblastoma multiforme, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma, oligodendroglioma, and ependymoma and related paraventricular mass lesions, neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other parenchymal tumors, including primary brain lymphoma, germ cell tumors, and pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor (malignant schwannoma), and neurocutaneous syndromes (phakomatoses), including neurofibromotosis, including Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous sclerosis, and Von Hippel-Lindau disease.

**[00326]** Cardiovascular disorders include, but are not limited to, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and

10085233.022802

rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia, disorders involving cardiac transplantation, and congestive heart failure.

**[00327]** Disorders involving blood vessels include, but are not limited to, responses of vascular cell walls to injury, such as endothelial dysfunction and endothelial activation and intimal thickening; vascular diseases including, but not limited to, congenital anomalies, such as arteriovenous fistula, atherosclerosis, and hypertensive vascular disease, such as hypertension; inflammatory disease--the vasculitides, such as giant cell (temporal) arteritis, Takayasu arteritis, polyarteritis nodosa (classic), Kawasaki syndrome (mucocutaneous lymph node syndrome), microscopic polyanglitis (microscopic polyarteritis, hypersensitivity or leukocytoclastic anglitis), Wegener granulomatosis, thromboanglitis obliterans (Buerger disease), vasculitis associated with other disorders, and infectious arteritis; Raynaud disease; aneurysms and dissection, such as abdominal aortic aneurysms, syphilitic (luetic) aneurysms, and aortic dissection (dissecting hematoma); disorders of veins and lymphatics, such as varicose veins, thrombophlebitis and phlebothrombosis, obstruction of superior vena cava (superior vena cava syndrome), obstruction of inferior vena cava (inferior vena cava syndrome), and lymphangitis and lymphedema; tumors, including benign tumors and tumor-like conditions, such as hemangioma, lymphangioma, glomus tumor (glomangioma), vascular ectasias, and bacillary angiomatosis, and intermediate-grade (borderline low-grade malignant) tumors, such as Kaposi's sarcoma and hemangloendothelioma, and malignant tumors, such as angiosarcoma and hemangiopericytoma; and pathology of therapeutic interventions in vascular disease, such as balloon angioplasty and related techniques and vascular replacement, such as coronary artery bypass graft surgery.

10085233.022802

**[00328]** Additionally, 93870 molecules can play an important role in the etiology of certain viral diseases (e.g., liver diseases), including but not limited to Hepatitis B, Hepatitis C and Herpes Simplex Virus (HSV). Modulators of 93870 activity can be used to control viral diseases. For example, the 93870 molecules can play a role in mediating viral protease

activities important for viral infection. The modulators can be used in the treatment and/or diagnosis of viral infected tissue or virus-associated tissue fibrosis, especially liver and liver fibrosis. Also, 93870 modulators can be used in the treatment and/or diagnosis of virus-associated carcinoma, especially hepatocellular cancer.

**[00329]** Hepatic disorders which can be treated or diagnosed by methods described herein include, but are not limited to, disorders associated with an accumulation in the liver of fibrous tissue, such as that resulting from an imbalance between production and degradation of the extracellular matrix accompanied by the collapse and condensation of preexisting fibers. The methods described herein can be used to diagnose or treat hepatocellular necrosis or injury induced by a wide variety of agents including processes which disturb homeostasis, such as an inflammatory process, tissue damage resulting from toxic injury or altered hepatic blood flow, and infections (e.g., bacterial, viral and parasitic). For example, the methods can be used for the early detection of hepatic injury, such as portal hypertension or hepatic fibrosis. In addition, the methods can be employed to detect liver fibrosis attributed to inborn errors of metabolism, for example, fibrosis resulting from a storage disorder such as Gaucher's disease (lipid abnormalities) or a glycogen storage disease, A1-antitrypsin deficiency; a disorder mediating the accumulation (e.g., storage) of an exogenous substance, for example, hemochromatosis (iron-overload syndrome) and copper storage diseases (Wilson's disease), disorders resulting in the accumulation of a toxic metabolite (e.g., tyrosinemia, fructosemia and galactosemia) and peroxisomal disorders (e.g., Zellweger syndrome). Additionally, the methods described herein may be useful for the early detection and treatment of liver injury associated with the administration of various chemicals or drugs, such as for example, methotrexate, isonizaid, oxyphenisatin, methyldopa, chlorpromazine, tolbutamide or alcohol, or which represents a hepatic manifestation of a vascular disorder such as obstruction of either the intrahepatic or extrahepatic bile flow or an alteration in hepatic circulation resulting, for example, from chronic heart failure, veno-occlusive disease, portal vein thrombosis or Budd-Chiari syndrome.

**[00330]** Additionally, 93870 may play an important role in the regulation of metabolism or pain disorders. Diseases of metabolic imbalance include, but are not limited to, obesity, anorexia nervosa, bullemia, cachexia, lipid disorders, and diabetes. Examples of pain disorders include, but are not limited to, pain response elicited during various forms of tissue injury, e.g., inflammation, infection, and ischemia, usually referred to as hyperalgesia (described in, for example, Fields, H.L., (1987) *Pain*, New York:McGraw-Hill); pain



10085233.022802

associated with muscoloskeletal disorders, e.g., joint pain; tooth pain; headaches; pain associated with surgery; pain related to irritable bowel syndrome; and chest pain.

Pharmacogenomics

**[00331]** The 93870 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on 93870 activity (e.g., 93870 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) 93870-associated disorders associated with aberrant or unwanted 93870 activity (e.g., disorders associated with hematopoiesis and immune disorders). In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) can be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician can consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 93870 molecule or 93870 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 93870 molecule or 93870 modulator.

**[00332]** Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons (e.g., Eichelbaum *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23:983-985; Linder *et al.*, (1997) *Clin. Chem.* 43:254-266). In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

**[00333]** One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association," relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the

human genome, each of which has two variants). Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high-resolution map can be generated from a combination of some ten million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP can be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that can be common among such genetically similar individuals.

10085233.022802

**[00334]** Alternatively, a method termed the "candidate gene approach" can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (e.g., a 93870 protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

**[00335]** Alternatively, a method termed "gene expression profiling," can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a 93870 molecule or 93870 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

**[00336]** Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 93870 molecule or 93870 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

**[00337]** The present invention further provides methods for identifying new agents, or combinations, that are based on identifying agents that modulate the activity of one or more of the gene products encoded by one or more of the 93870 genes of the present invention, wherein these products can be associated with resistance of the cells to a therapeutic agent.

Specifically, the activity of the proteins encoded by the 93870 genes of the present invention can be used as a basis for identifying agents for overcoming agent resistance. By blocking the activity of one or more of the resistance proteins, target cells, e.g., hematopoietic cells, will become sensitive to treatment with an agent that the unmodified target cells were resistant to.

**[00338]** Monitoring the influence of agents (e.g., drugs) on the expression or activity of a 93870 protein can be applied in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 93870 gene expression, protein levels, or up-regulate 93870 activity, can be monitored in clinical trials of subjects exhibiting decreased 93870 gene expression, protein levels, or down-regulated 93870 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 93870 gene expression, protein levels, or down-regulate 93870 activity, can be monitored in clinical trials of subjects exhibiting increased 93870 gene expression, protein levels, or up-regulated 93870 activity. In such clinical trials, the expression or activity of a 93870 gene, and preferably, other genes that have been implicated in, for example, a 93870-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

Other Embodiments

**[00339]** In another aspect, the invention features a method of analyzing a plurality of capture probes. The method is useful, e.g., to analyze gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., a nucleic acid or peptide sequence, wherein the capture probes are from a cell or subject which expresses 93870 or from a cell or subject in which a 93870 mediated response has been elicited; contacting the array with a 93870 nucleic acid (preferably purified), a 93870 polypeptide (preferably purified), or an anti-93870 antibody, and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by a signal generated from a label attached to the 93870 nucleic acid, polypeptide, or antibody.

10085233.022802

**[00340]** The capture probes can be a set of nucleic acids from a selected sample, e.g., a sample of nucleic acids derived from a control or non-stimulated tissue or cell.

**[00341]** The method can include contacting the 93870 nucleic acid, polypeptide, or antibody with a first array having a plurality of capture probes and a second array having a different plurality of capture probes. The results of each hybridization can be compared, e.g., to analyze differences in expression between a first and second sample. The first plurality of capture probes can be from a control sample, e.g., a wild type, normal, or non-diseased, non-stimulated, sample, e.g., a biological fluid, tissue, or cell sample. The second plurality of capture probes can be from an experimental sample, e.g., a mutant type, at risk, disease-state or disorder-state, or stimulated, sample, e.g., a biological fluid, tissue, or cell sample.

**[00342]** The plurality of capture probes can be a plurality of nucleic acid probes each of which specifically hybridizes, with an allele of 93870. Such methods can be used to diagnose a subject, e.g., to evaluate risk for a disease or disorder, to evaluate suitability of a selected treatment for a subject, to evaluate whether a subject has a disease or disorder.

**[00343]** The method can be used to detect SNPs, as described above.

**[00344]** In another aspect, the invention features, a method of analyzing 93870, e.g., analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a 93870 nucleic acid or amino acid sequence; comparing the 93870 sequence with one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database; to thereby analyze 93870.

**[00345]** The method can include evaluating the sequence identity between a 93870 sequence and a database sequence. The method can be performed by accessing the database at a second site, e.g., over the internet. Preferred databases include GenBank™ and SwissProt.

**[00346]** In another aspect, the invention features, a set of oligonucleotides, useful, e.g., for identifying SNP's, or identifying specific alleles of 93870. The set includes a plurality of oligonucleotides, each of which has a different nucleotide at an interrogation position, e.g., an SNP or the site of a mutation. In a preferred embodiment, the oligonucleotides of the plurality identical in sequence with one another (except for differences in length). The oligonucleotides can be provided with differential labels, such that an oligonucleotides which hybridizes to one allele provides a signal that is distinguishable from an oligonucleotides which hybridizes to a second allele.



10085233.022802

**[00347]** The sequence of a 93870 molecules is provided in a variety of mediums to facilitate use thereof. A sequence can be provided as a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a 93870 molecule. Such a manufacture can provide a nucleotide or amino acid sequence, e.g., an open reading frame, in a form which allows examination of the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exists in nature or in purified form.

**[00348]** A 93870 nucleotide or amino acid sequence can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact disc and CD-ROM; electrical storage media such as RAM, ROM, EPROM, EEPROM, and the like; and general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having thereon 93870 sequence information of the present invention.

**[00349]** As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus of other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as personal digital assistants (PDAs), cellular phones, pagers, and the like; and local and distributed processing systems.

**[00350]** As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the 93870 sequence information.

**[00351]** A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a 93870 nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software

such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of data processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

**[00352]** By providing the 93870 nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. A search is used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

**[00353]** The present invention therefore provides a medium for holding instructions for performing a method for determining whether a subject has a GPCR or 93870-associated disease or disorder or a pre-disposition to a GPCR or 93870-associated disease or disorder, wherein the method comprises the steps of determining 93870 sequence information associated with the subject and based on the 93870 sequence information, determining whether the subject has a GPCR or 93870-associated disease or disorder and/or recommending a particular treatment for the disease, disorder, or pre-disease condition.

**[00354]** The present invention further provides in an electronic system and/or in a network, a method for determining whether a subject has a 93870 or GPCR-associated disease or disorder or a pre-disposition to a disease associated with 93870, wherein the method comprises the steps of determining 93870 sequence information associated with the subject, and based on the 93870 sequence information, determining whether the subject has a GPCR or 93870-associated disease or disorder or a pre-disposition to a GPCR or 93870-associated disease or disorder, and/or recommending a particular treatment for the disease, disorder, or pre-disease condition. The method may further comprise the step of receiving phenotypic information associated with the subject and/or acquiring from a network phenotypic information associated with the subject.

**[00355]** The present invention also provides in a network, a method for determining whether a subject has a GPCR or 93870-associated disease or disorder or a pre-disposition to a GPCR or 93870-associated disease or disorder, said method comprising the steps of receiving 93870 sequence information from the subject and/or information related thereto,

receiving phenotypic information associated with the subject, acquiring information from the network corresponding to 93870 and/or corresponding to a GPCR or 93870-associated disease or disorder, and based on one or more of the phenotypic information, the 93870 information (e.g., sequence information and/or information related thereto), and the acquired information, determining whether the subject has a GPCR or 93870-associated disease or disorder or a pre-disposition to a GPCR or 93870-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder, or pre-disease condition.

**[00356]** The present invention also provides a business method for determining whether a subject has a GPCR or 93870-associated disease or disorder or a pre-disposition to a GPCR or 93870-associated disease or disorder, said method comprising the steps of receiving information related to 93870 (e.g., sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to 93870 and/or related to a GPCR or 93870-associated disease or disorder, and based on one or more of the phenotypic information, the 93870 information, and the acquired information, determining whether the subject has a GPCR or 93870-associated disease or disorder or a pre-disposition to a GPCR or 93870-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder, or pre-disease condition.

10085233.022802

**[00357]** The invention also includes an array comprising a 93870 sequence of the present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression, one of which can be 93870. This allows a profile to be developed showing a battery of genes specifically expressed in one or more tissues.

**[00358]** In addition to such qualitative information, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue if ascertainable. Thus, genes can be grouped on the basis of their tissue expression *per se* and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression in that tissue. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. In this context, the effect of one cell type on another cell type in

response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

**[00359]** In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of a 93870 -associated disease or disorder, progression of 93870 -associated disease or disorder, and processes, such a cellular transformation associated with the 93870 -associated disease or disorder.

**[00360]** The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells (*e.g.*, acertaining the effect of 93870 expression on the expression of other genes). This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

**[00361]** The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes (*e.g.*, including 93870) that could serve as a molecular target for diagnosis or therapeutic intervention.

**[00362]** As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. Typical sequence lengths of a target sequence are from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

**[00363]** Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the



10085233.022802

computer-based systems of the present invention. Examples of such software include, but are not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBI).

**[00364]** Thus, the invention features a method of making a computer readable record of a sequence of a 93870 sequence which includes recording the sequence on a computer readable matrix. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

**[00365]** In another aspect, the invention features, a method of analyzing a sequence. The method includes: providing a 93870 sequence, or record, in computer readable form; comparing a second sequence to the 93870 sequence; thereby analyzing a sequence. Comparison can include comparing to sequences for sequence identity or determining if one sequence is included within the other, e.g., determining if the 93870 sequence includes a sequence being compared. In a preferred embodiment the 93870 or second sequence is stored on a first computer, e.g., at a first site and the comparison is performed, read, or recorded on a second computer, e.g., at a second site. E.g., the 93870 or second sequence can be stored in a public or proprietary database in one computer, and the results of the comparison performed, read, or recorded on a second computer. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

**[00366]** The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

## EXEMPLIFICATION

Gene Expression Analysis

**[00366]** Total RNA was prepared from various human tissues by a single step extraction method using RNA STAT-60 according to the manufacturer's instructions (TelTest, Inc). Each RNA preparation was treated with DNase I (Ambion) at 37°C for 1 hour. DNAse I



10085232.020802

treatment was determined to be complete if the sample required at least 38 PCR amplification cycles to reach a threshold level of fluorescence using β-2 microglobulin as an internal amplicon reference. The integrity of the RNA samples following DNase I treatment was confirmed by agarose gel electrophoresis and ethidium bromide staining. After phenol extraction cDNA was prepared from the sample using the SUPERSCRIPT™ Choice System following the manufacturer's instructions (GibcoBRL). A negative control of RNA without reverse transcriptase was mock reverse transcribed for each RNA sample.

**[00367]** Human 93870 expression was measured by TaqMan® quantitative PCR (Perkin Elmer Applied Biosystems) in cDNA prepared from a variety of normal and diseased (e.g., cancerous) human tissues or cell lines.

**[00368]** Probes were designed by PrimerExpress software (PE Biosystems) based on the sequence of the human 93870 genes. Each human 93870 gene probe was labeled using FAM (6-carboxyfluorescein), and the β2-microglobulin reference probe was labeled with a different fluorescent dye, VIC. The differential labeling of the target gene and internal reference gene thus enabled measurement in same well. Forward and reverse primers and the probes for both β2-microglobulin and target gene were added to the TaqMan® Universal PCR Master Mix (PE Applied Biosystems). Although the final concentration of primer and probe could vary, each was internally consistent within a given experiment. A typical experiment contained 200nM of forward and reverse primers plus 100nM probe for β-2 microglobulin and 600 nM forward and reverse primers plus 200 nM probe for the target gene. TaqMan matrix experiments were carried out on an ABI PRISM 7700 Sequence Detection System (PE Applied Biosystems). The thermal cycler conditions were as follows: hold for 2 min at 50°C and 10 min at 95°C, followed by two-step PCR for 40 cycles of 95°C for 15 sec followed by 60°C for 1 min.

**[00369]** The following method was used to quantitatively calculate human 93870 gene expression in the various tissues relative to β-2 microglobulin expression in the same tissue. The threshold cycle (Ct) value is defined as the cycle at which a statistically significant increase in fluorescence is detected. A lower Ct value is indicative of a higher mRNA concentration. The Ct value of the human 93870 genes is normalized by subtracting the Ct value of the β-2 microglobulin gene to obtain a $_{\Delta}Ct$ value using the following formula: $_{\Delta}Ct = Ct_{\text{human 93870}} - Ct_{\beta\text{-2 microglobulin}}$. Expression is then calibrated against a cDNA sample showing a comparatively low level of expression of the human 93870 gene. The $_{\Delta}Ct$ value for the calibrator sample is then subtracted from $_{\Delta}Ct$ for each tissue sample according to the

following formula: ${}_{\Delta\Delta}Ct={}_{\Delta}Ct\text{-}_{sample} - {}_{\Delta}Ct\text{-}_{calibrator}$. Relative expression is then calculated using the arithmetic formula given by $2^{-\Delta\Delta Ct}$. Expression of the target human 93870 genes in each of the tissues tested is then graphically represented as discussed in more detail below.

**[00370]** The results indicate high levels of 93870 expression in normal bone marrow mononuclear cells and neutrophils; moderate levels of 93870 expression in human osteoblasts; and 93870 expression in megakaryocytes.

**[00371]** The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

**Equivalents**

**[00372]** Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.